PCT WELTORGANISATION FÜR GEISTIGES EIGENTUM
Internationales Büro
INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

(51) Internationale Patentklassifikation 6 : C12N 15/54, 9/12, 15/11, 15/85, C12Q 1/68, A01K 67/027

A2

(11) Internationale Veröffentlichungsnummer: **WO 99/33998**

(43) Internationales Veröffentlichungsdatum: 8. Juli 1999 (08.07.99)

(21) Internationales Aktenzeichen: PCT/EP98/08216

(22) Internationales Anmeldedatum: 22. Dezember 1998 (22.12.98)

(30) Prioritätsdaten:
197 57 984.1 24. Dezember 1997 (24.12.97) DE

(71) Anmelder *(für alle Bestimmungsstaaten ausser US)*: BAYER AKTIENGESELLSCHAFT [DE/DE]; D–51368 Leverkusen (DE).

(72) Erfinder; und

(75) Erfinder/Anmelder *(nur für US)*: HAGEN, Gustav [DE/DE]; Bertha–von–Suttner–Strasse 31, D–51373 Leverkusen (DE). WICK, Maresa [DE/DE]; Andreas–Gryphius–Strasse 26, D–51065 Köln (DE). ZUBOV, Dmitry [RU/DE]; Roggendorfstrasse 59, D–51061 Köln (DE).

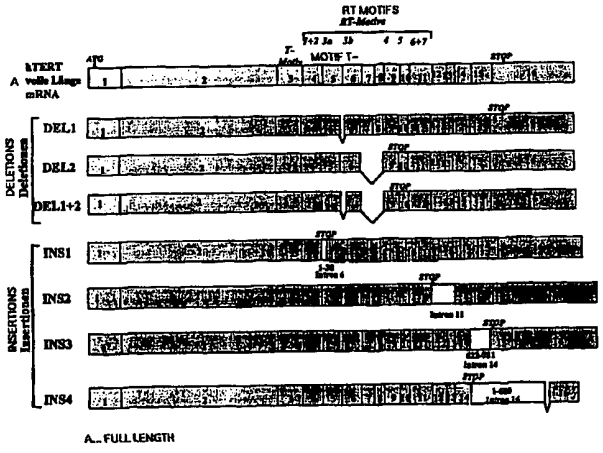
(74) Gemeinsamer Vertreter: BAYER AKTIENGESELLSCHAFT; D–51368 Leverkusen (DE).

(81) Bestimmungsstaaten: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht
Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.

(54) Title: REGULATORY DNA SEQUENCES OF THE HUMAN CATALYTIC TELOMERASE SUB–UNIT GENE, DIAGNOSTIC AND THERAPEUTIC USE THEREOF

(54) Bezeichnung: REGULATORISCHE DNA–SEQUENZEN DES GENS DER HUMANEN KATALYTISCHEN TELOMERASE–UNTEREINHEIT UND DEREN DIAGNOSTISCHE UND THERAPEUTISCHE VERWENDUNG



(57) Abstract

The present invention relates to regulatory DNA sequences containing promotor sequences, in addition to intervening sequences, for the human catalytic telomerase sub–unit gene. The invention also relates to the use of said DNA sequences for pharmaceutical, diagnostic and therapeutic purposes, especially in the treatment of cancer and ageing.

(57) **Zusammenfassung**

Diese Erfindung betrifft regulatorische DNA–Sequenzen, beinhaltend Promotorsequenzen, sowie Intronsequenzen, für das Gen der humanen kalytischen Telomerase–Untereinheit. Darüber hinaus betrifft diese Erfindung die Verwendung dieser DNA–Sequenzen für pharmazeutische, diagnostische und therapeutische Zwecke, vor allem in der Behandlung von Krebs und Alterung.

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidschan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland			TR	Türkei
BG	Bulgarien	HU	Ungarn	ML	Mali	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MN	Mongolei	UA	Ukraine
BR	Brasilien	IL	Israel	MR	Mauretanien	UG	Uganda
BY	Belarus	IS	Island	MW	Malawi	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	MX	Mexiko		
CF	Zentralafrikanische Republik	JP	Japan	NE	Niger	UZ	Usbekistan
CG	Kongo	KE	Kenia	NL	Niederlande	VN	Vietnam
CH	Schweiz	KG	Kirgisistan	NO	Norwegen	YU	Jugoslawien
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	NZ	Neuseeland	ZW	Zimbabwe
CM	Kamerun			PL	Polen		
CN	China	KR	Republik Korea	PT	Portugal		
CU	Kuba	KZ	Kasachstan	RO	Rumänien		
CZ	Tschechische Republik	LC	St. Lucia	RU	Russische Föderation		
DE	Deutschland	LI	Liechtenstein	SD	Sudan		
DK	Dänemark	LK	Sri Lanka	SE	Schweden		
EE	Estland	LR	Liberia	SG	Singapur		

Regulatorische DNA-Sequenzen des Gens der humanen katalytischen Telomerase-Untereinheit und deren diagnostische und therapeutische Verwendung

Aufbau und Funktion der Chromosomenenden

Das genetische Material eukaryontischer Zellen ist auf linearen Chromosomen verteilt. Die Enden der Erbanlagen werden, abgeleitet von den griechischen Wörtern *telos* (Ende) und *meros* (Teil, Segment), als Telomere bezeichnet. Die meisten Telomere bestehen aus Wiederholungen von kurzen Sequenzen, die überwiegend aus Thymin und Guanin aufgebaut sind (Zakian, 1995). In allen bislang untersuchten Wirbeltieren werden die Telomere aus der Sequenz TTAGGG aufgebaut (Meyne *et al.*, 1989).

Die Telomere üben verschiedene wichtige Funktionen aus. Sie verhindern die Fusion von Chromosomen (McClintock, 1941) und damit die Entstehung von dizentrischen Erbanlagen. Solche Chromosomen mit zwei Centromeren können durch Verlust der Heterozygotie bzw. Verdopplung oder Verlust von Genen zur Entwicklung von Krebs führen.

Desweiteren dienen Telomere dazu, intakte Erbanlagen von beschädigten zu unterscheiden. So stellten Hefezellen ihre Zellteilung ein, wenn sie ein Chromosom ohne Telomer enthielten (Sandell und Zakian, 1993).

Eine weitere wichtige Aufgabe erfüllen Telomere bei der DNA-Replikation eukaryontischer Zellen. Im Gegensatz zu den zirkulären Genomen von Prokaryonten können die linearen Chromosomen der Eukaryonten von dem DNA Polymerase-Komplex nicht vollständig repliziert werden. Zur Initiation der DNA-Replikation sind RNA-Primer notwendig. Nach Abspaltung der RNA-Primer, Verlängerung der Okazaki-Fragmente und anschließender Ligation fehlt dem neu-synthetisierten DNA-Strang das 5'-Ende, denn dort kann der RNA-Primer nicht durch DNA ersetzt

werden. Ohne besondere Schutzmechanismen würden daher die Chromosomen mit jeder Zellteilung schrumpfen ("end-replication problem"; Harley *et al.*, 1990). Die nicht-kodierenden Telomersequenzen stellen vermutlich eine Pufferzone dar, um dem Verlust von Genen vorzubeugen (Sandell und Zakian, 1993).

Darüberhinaus spielen Telomere auch eine wichtige Rolle bei der Regulation der zellulären Alterung (Olovnikov, 1973). Humane somatische Zellen zeigen in Kultur eine limitierte Replikationskapazität; sie werden nach einer gewissen Zeit seneszent. In diesem Zustand teilen sich die Zellen selbst nach Stimulierung mit Wachstumsfaktoren nicht mehr, sterben aber nicht, sondern bleiben metabolisch aktiv (Goldstein, 1990). Verschiedene Beobachtungen sprechen für die Hypothese, daß eine Zelle anhand der Länge ihrer Telomere bestimmt, wie oft sie sich noch teilen kann (Allsopp *et al.*, 1992).

Zusammenfassend besitzen die Telomere somit zentrale Funktionen bei der Alterung von Zellen sowie der Stabilisierung des genetischen Materials und Verhinderung von Krebs.

Das Enzym Telomerase synthetisiert die Telomere

Wie oben beschrieben können Organismen mit linearen Chromosomen ohne einen speziellen Schutzmechanismus ihr Genom nur unvollständig replizieren. Die meisten Eukaryonten verwenden zur Regeneration der Telomersequenzen ein spezielles Enzym, die Telomerase. In den bislang untersuchten Einzellern wird Telomerase konstitutiv exprimiert. Dagegen wurde in Menschen die Telomerase-Aktivität nur in Keimzellen und Tumorzellen gemessen, wogegen benachbartes somatisches Gewebe keine Telomerase enthielt (Kim *et al.*, 1994).

Funktionell kann die Telomerase auch als terminale Telomertransferase bezeichnet werden, die als Multiproteinkomplex im Zellkern lokalisiert ist. Während der RNA-Anteil der humanen Telomerase schon seit längerem bekannt ist (Feng *et al.*, 1995),

wurde kürzlich die katalytische Untereinheit dieser Enzymgruppe in verschiedenen Organismen identifiziert (Lingner *et al.*, 1997; vgl. unsere ebenfalls anhängige Anmeldung PCT EP/98/03468). Diese katalytischen Untereinheiten der Telomerase sind sowohl untereinander als auch zu bisher allen bekannten reversen Transkriptasen auffällig homolog.

Auch in WO 98/14592 werden Nukleinsäure- und Aminosäuresequenzen der katalytischen Telomerase-Untereinheit beschrieben.

Aktivierung der Telomerase in menschlichen Tumoren

Eine Aktivität der Telomerase konnte in Menschen ursprünglich nur in Keimbahnzellen, nicht aber in normalen somatischen Zellen (Hastie *et al.*, 1990; Kim *et al.*, 1994) nachgewiesen werden. Nach der Entwicklung eines sensitiveren Nachweisverfahrens (Kim *et al.*, 1994) wurde auch in hematopoietischen Zellen eine geringe Telomeraseaktivität detektiert (Broccoli *et al.*, 1995; Counter *et al.*, 1995; Hiyama *et al.*, 1995). Allerdings wiesen diese Zellen trotzdem eine Reduktion der Telomere auf (Vaziri *et al.*, 1994; Counter *et al.*, 1995). Noch ist nicht geklärt, ob die Menge an Enzym in diesen Zellen nicht ausreichend für eine Kompensation des Telomerverlustes ist, oder ob die gemessene Telomerase-Aktivität von einer Subpopulation, z.B. unvollständig ausdifferenzierten $CD34^{+}38^{+}$-Vorläuferzellen, herrührt (Hiyama *et al.*, 1995). Zur Klärung wäre ein Nachweis der Telomerase-Aktivität in einer einzelnen Zelle nötig.

Interessanterweise wurde jedoch in einer großen Zahl der bislang getesteten Tumorgewebe eine signifikante Telomerase-Aktivität nachgewiesen (1734/2031, 85 %; Shay, 1997), während in normalem somatischen Gewebe keine Aktivität gefunden wurde (1/196, <1 %, Shay, 1997). Verschiedene Untersuchungen zeigten außerdem, daß in seneszenten Zellen, die mit viralen Oncoproteinen transformiert wurden, die Telomere weiterhin schrumpften und Telomerase nur in der Subpopulation entdeckt werden konnte, die die Wachstumskrise überlebte (Counter *et al.*, 1992). In diesen immortalisierten Zellen waren auch die Telomere stabil (Counter *et al.*, 1992). Ähnli-

che Befunde aus Untersuchungen an Mäusen (Blasco *et al.*, 1996) stützen die Annahme, daß eine Reaktivierung der Telomerase ein spätes Ereignis in der Tumorgenese ist.

Basierend auf diesen Ergebnissen wurde eine "Telomerase-Hypothese" entwickelt, die den Verlust von Telomersequenzen und Zellalterung mit der Aktivität von Telomerase und der Entstehung von Krebs verbindet. In langlebigen Spezies wie dem Menschen kann das Schrumpfen der Telomere als ein Mechanismus zur Tumorsuppression angesehen werden. Ausdifferenzierte Zellen, die keine Telomerase enthalten, stellen bei einer bestimmten Länge der Telomere ihre Zellteilung ein. Mutiert eine solche Zelle, so kann aus ihr nur dann ein Tumor entstehen, wenn die Zelle ihre Telomere verlängern kann. Ansonsten würde die Zelle weiterhin Telomersequenzen verlieren, bis ihre Chromosomen instabil werden und sie schließlich zugrunde geht. Die Reaktivierung der Telomerase ist vermutlich der Hauptmechanismus von Tumorzellen zur Stabilisation ihrer Telomere.

Aus diesen Beobachtungen und Überlegungen ergibt sich, daß eine Inhibition der Telomerase eine Therapie von Tumoren erlauben sollte. Konventionelle Krebstherapien mit Zytostatika oder kurzwelligen Strahlen schädigen nicht nur die Tumorzellen, sondern alle sich teilenden Zellen des Körpers. Da aber außer Tumorzellen nur Keimbahnzellen eine signifikante Telomerase-Aktivität enthalten, würden Telomerase-Inhibitoren spezifischer die Tumorzellen angreifen und somit weniger unerwünschte Nebenwirkungen hervorrufen. In allen bislang getesteten Tumorgeweben wurde eine Telomerase-Aktivität nachgewiesen, so daß diese Therapeutika gegen alle Krebsarten eingesetzt werden könnten. Die Wirkung von Telomerase-Inhibitoren würde dann eintreten, wenn die Telomere der Zellen sich soweit verkürzt haben, daß das Genom instabil wird. Da Tumorzellen meist kürzere Telomere aufweisen als normale somatische Zellen, würden zuerst Krebszellen durch Telomerase-Inhibitoren eliminiert werden. Zellen mit langen Telomeren, wie die Keimzellen, würden dagegen erst viel später geschädigt werden. Telomerase-Inhibitoren stellen somit einen zukunftsweisenden Weg für die Therapierung von Krebs dar.

Eindeutige Antworten auf die Frage nach der Art und den Angriffspunkten physiologischer Telomerase-Inhibitoren werden möglich sein, wenn auch die Regulation der Genexpression der Telomerase identifiziert ist.

Regulation der Genexpression in Eukaryonten

Die eukaryotische Genexpression, d.h. der zelluläre Informationsfluß von der DNA über die RNA zum Protein, weist vielfältige Ansatzpunkte für regulatorische Mechanismen auf. Einzelne Kontrollstufen sind z.B. die Gen-Amplifikation, Rekombination von Genloci, Chromatinstruktur, DNA-Methylierung, Transkription, posttranskriptionelle mRNA-Modifikationen, mRNA-Transport, Translation und post-translationale Proteinmodifikationen. Nach bisherigen Studien besitzt die Kontrolle auf der Ebene der Transkriptionsinitiation die größte Bedeutung (Latchman, 1991).

Unmittelbar stromaufwärts vom Transkriptionsstart eines von der RNA-Polymerase II transkribierten Gens liegt eine Region, die für die Steuerung der Transkription verantwortlich ist und als Promotorregion bezeichnet wird. Ein Vergleich der Nukleotidsequenzen von Promotorregionen vieler bekannter Gene zeigt, daß bestimmte Sequenzmotive in dieser Region häufig vorkommen. Zu diesen Elementen gehören unter anderem die TATA-Box, die CCAAT-Box und die GC-Box, die von spezifischen Proteinen erkannt werden. Die TATA-Box, die etwa 30 Nukleotide stromaufwärts vom Transkriptionsstart entfernt positioniert ist, wird z.B. von der TFIID-Untereinheit TBP („TATA-box binding protein") erkannt, wogegen bestimmte GC-reiche Sequenzabschnitte vom Transkriptionsfaktor Sp1 („specificity protein1") spezifisch gebunden werden.

Funktionell kann man den Promotor in einen regulativen und einen konstitutiven Abschnitt unterteilen (Latchman, 1991). Der konstitutive Kontrollbereich umfaßt den sogenannten Kernpromotor („corepromoter"), der die korrekte Initiation der Transkription ermöglicht. Er enthält die als UPE's (upstream promoter elements") be-

schriebenen Sequenzelemente, die für eine effiziente Transkription notwendig sind. Die regulativen Kontrollabschnitte, die mit den UPE's verflochten sein können, weisen Sequenzelemente auf, die an der signalabhängigen Regulation der Transkription durch Hormone, Wachstumsfaktoren usw. beteiligt sein können. Sie vermitteln gewebs- oder zellspezifische Promotoreigenschaften.

Ein charakteristisches Merkmal eukaryotischer Gene sind DNA-Abschnitte, die über vergleichsweise große Distanzen hinweg Einfluß auf die Genexpression nehmen können. Diese Elemente können stromaufwärts, stromabwärts oder innerhalb einer Transkriptionseinheit lokalisiert sein und unabhängig von ihrer Orientierung ihre Funktion wahrnehmen. Diese Sequenzabschnitte können die Promotoraktivität verstärken (Enhancer) oder abschwächen (Silencer). Ähnlich wie die Promotorregionen beherbergen auch Enhancer und Silencer mehrere Bindungsstellen für Transkriptionsfaktoren.

Die Erfindung betrifft die DNA-Sequenzen aus der 5'-flankierenden Region des Gens der katalytisch aktiven humanen Telomerase-Untereinheit sowie Intron-Sequenzen für dieses Gen.

Die Erfindung betrifft insbesondere die 5'-flankierende regulatorische DNA-Sequenz, enthaltend die Promotor-DNA-Sequenz für das Gen der humanen katalytischen Telomerase Untereinheit gemäß Fig. 10 (SEQ ID NO 3).

Die Erfindung betrifft weiterhin regulatorisch wirksame Teilbereiche der 5'-flankierenden regulatorischen DNA-Sequenz gemäß Fig. 4 (SEQ ID NO 1).

Weiterhin sind Gegenstand der vorliegenden Erfindung Intron-Sequenzen für das Gen der humanen katalytischen Telomerase-Untereinheit, insbesondere solche, die regulatorische Wirkung haben. Die erfindungsgemäßen Intronsequenzen werden im Rahmen von Beispiel 5 detailliert beschrieben (vgl. SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 und 20).

Die Erfindung betrifft weiterhin ein rekombinantes Konstrukt, das die erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierende DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereiche davon beinhaltet.

Bevorzugt sind rekombinante Konstrukte, die neben den erfindungsgemäßen DNA-Sequenzen, insbesondere der 5'-flankierenden DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereichen davon, eine oder mehrere weitere DNA-Sequenzen, die für Polypeptide oder Proteine kodieren, enthalten.

Gemäß einer besonders bevorzugten Ausführungsform kodieren diese weiteren DNA-Sequenzen für antitumorale Proteine.

Besonders bevorzugte antitumorale Proteine sind solche, die die Angiogenese direkt oder indirekt inhibieren. Zu diesen Proteinen zählen beispielsweise:

Plasminogenaktivatorinhibitor (PAI-1), PAI-2, PAI-3, Angiostatin, Endostatin, Platelet factor 4, TIMP-1, TIMP-2, TIMP-3, Leukemia Inhibitory Factor (LIF).

Ebenfalls besonders bevorzugt sind antitumorale Proteine, welche direkt oder indirekt eine zytostatische Wirkung auf Tumoren aufweisen. Hierzu zählen im besonderen:

Perforin, Granzym, IL-2, IL-4, IL-12, Interferone, wie beispielsweise IFN-α, IFN-ß, IFN-γ, TNF, TNF-α, TNF-ß, Oncostatin M; Tumorsuppressorgene, wie z.B. p53, Retinoblastoma.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche gegebenenfalls zusätzlich zur antitumoralen Wirkung Entzündungen stimulieren und hierdurch zur Elimination von Tumorzellen beitragen. Hierzu zählen beispielsweise:

RANTES, Monocyte chemotactic and activating factor (MCAF), IL-8, Macrophage inflammatory protein (MIP-1α,-ß), Neutrophil activating protein-2 (NAP-2), IL-3, IL-5, human leukemia inhibitory factor (LIF), IL-7, IL-11, IL-13, GM-CSF, G-CSF, M-CSF.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche aufgrund ihrer Wirkung als Enzyme in der Lage sind, Vorstufen eines antitumoralen Wirkstoffes in einen antitumoralen Wirkstoff zu überführen. Zu diesen Enzymen zählen beispielsweise:

Herpes Simplex Virus Thymidinkinase, Varizella Zoster Virus Thymidinkinase, bakterielle Nitroreductase, bakterielle ß- Glukuronidase, pflanzliche ß-Glukuronidase aus *Secale careale*, humane Glukuronidase, humane Carboxypeptidase, bakterielle Carboxypeptidase, bakterielle ß-Lactamase, bakterielle Cytosindeaminidase, humane Katalase bzw. Phosphatase, humane alkalische Phosphatase, Typ 5 saure Phosphatase, humane Lysooxidase, humane saure D-Aminooxidase, humane Glutathion Peroxidase, humane Eosinophilen Peroxidase, humane Schilddrüsen Peroxidase.

Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für Faktor VIII, IX oder Teilfragmente davon kodieren. Zu diesen DNA-Sequenzen zählen auch andere Blutgerinnungsfaktoren

Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für ein Reporterprotein kodieren. Zu diesen Reporterproteinen zählen beispielsweise:

Chloramphenicolacetyltransferase (CAT), Glühwürmchen Luziferase (LUC), ß-Galaktosidase (ß-Gal), Sezernierte alkalische Phosphatase (SEAP), Humanes Wachstumshormon (hGH), ß-Glukuronidase (GUS), Grün-fluoreszierendes Protein (GFP) und alle davon abgeleiteten Varianten, Aquarin, Obelin.

Erfindungsgemäße rekombinante Konstrukte können auch DNA kodierend für die humane katalytische Telomerase Untereinheit und deren Varianten und Fragmente in antisense Orientierung enthalten. Gegebenenfalls können diese Konstrukte auch andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente in antisense Orientierung enthalten.

Die rekombinanten Konstrukte können neben der DNA, kodierend für die humane katalytische Telomerase Untereinheit, sowie deren Varianten und Fragmente auch andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente enthalten.

Die Erfindung betrifft weiterhin einen Vektor, enthaltend die oben genannten erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierenden DNA-Sequenzen, sowie eine oder mehrere der oben genannten anderen DNA-Sequenzen.

Bevorzugter Vektor für solche Konstrukte ist ein Virus, beispielsweise ein Retrovirus, Adenovirus, adeno-assoziiertes Virus, Herpes Simplex Virus, Vaccina Virus, lentivirales Virus, Sindbis Virus und ein Semliki Forest Virus.

Ebenfalls bevorzugt sind Plasmide als Vektoren.

Die Erfindung betrifft weiterhin pharmazeutische Präparate, enthaltend erfindungsgemäße rekombinante Konstrukte bzw. Vektoren; beispielsweise eine Zubereitung in einem kolloidalen Dispersionssystem.

Geeignete kolloidale Dispersionssysteme sind beispielsweise Liposome oder Polylysin-Liganden.

Die Zubereitungen der erfindungsgemäßen Konstrukte bzw. Vektoren in kolloidalen Dispersionssystemen können um einen Liganden ergänzt sein, der an Membranstruk-

turen von Tumorzellen bindet. Ein solcher Ligand kann z.B. an das Konstrukt bzw. den Vektor angeknüpft sein oder auch Bestandteil der Liposomenstruktur sein.

Geeignete Liganden sind insbesondere polyklonale oder monoklonale Antikörper oder Antikörperfragmente hiervon, die mit ihren variablen Domänen an Membranstrukturen von Tumorzellen binden, oder endständige Mannose-tragende Substanzen, Zytokine, Wachstumsfaktoren oder Fragmente bzw. Teilsequenzen hiervon, die an Rezeptoren auf Tumorzellen binden.

Entsprechende Membranstrukturen sind beispielsweise Rezeptoren für ein Zytokin oder einen Wachstumsfaktor, wie z.B. IL-1, EGF, PDGF, VEGF, TGF ß, Insulin oder Insulin-like Growth Factor (ILGF), oder Adhäsionsmoleküle, wie z. B. SLeX, LFA-1, MAC-1, LECAM-1 oder VLA-4, oder der Mannose-6-Phosphat-Rezeptor.

Zur vorliegenden Erfindung gehören pharmazeutische Zubereitungen, die neben den erfindungsgemäßen Vektorkonstrukten auch nichttoxische, inerte, pharmazeutisch geeignete Trägerstoffe enthalten können. Vorstellbar sind die Applikation (z.B. intravenös, intraarteriell, intramuskulär, subkutan, intradermal, anal, vaginal, nasal, transdermal, intraperitonal, als Aerosol oder oral) am Ort eines Tumors oder die systemische Applikation dieser Zubereitungen.

Die erfindungsgemäßen Vektorkonstrukte können in der Gentherapie eingesetzt werden.

Die Erfindung betrifft weiterhin eine rekombinante Wirtszelle, insbesondere eine rekombinante eukaryotische Wirtszelle, enthaltend die vorstehend beschriebenen Konstrukte bzw. Vektoren.

Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit beeinflussen, wobei dieses Verfahren folgende Schritte umfaßt:

A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend die erfindungsgemäße regulatorische DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon, funktionell verknüpft mit einem Reportergen,

B. Messung des Substanzeffektes auf die Reportergenexpression.

Das Verfahren kann eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit verstärken.

Das Verfahren kann weiterhin eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit inhibieren.

Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Faktoren, die spezifisch an Fragmente der erfindungsgemäßen DNA-Fragmente, insbesondere der 5´-flankierenden regulatorischen DNA-Sequenz der katalytischen Telomerase Untereinheit, binden. Diese Methode beinhaltet ein Screening einer Expressions-cDNA-Bibliothek mit der vorstehend beschriebenen DNA-Sequenz oder Teilfragmenten unterschiedlichster Länge als Sonde.

Die vorstehend beschriebenen Konstrukte bzw. Vektoren können auch zur Herstellung transgener Tiere verwendet werden.

Die Erfindung betrifft weiterhin ein Verfahren zur Detektion Telomerase-assoziierter Zustände bei einem Patienten, das folgende Schritte umfaßt:

A. Inkubation eines Konstruktes bzw. Vektors, enthaltend die erfindungsgemäße DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon sowie ein Reportergen mit Körperflüssigkeiten oder zellulären Proben,

B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten;

C. Vergleich des diagnostischen Werts mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe;

Detektion diagnostischer Werte, die höher oder niedriger als Standardvergleichswerte liegen, indiziert einen Telomerase-assoziierten Zustand, der wiederum einen pathogenen Zustand indiziert.

Erläuterung der Abbildungen:

Fig. 1: Southern Blot-Analyse mit genomischer DNA verschiedener Spezies

A: Foto eines Ethidiumbromid gefärbten 0,7 %igen Agarosegels mit etwa 4 µg Eco RI geschnittener genomischer DNA. Die Spur 1 enthält Hind III geschnittene λ-DNA als Größenmarker (23,5, 9,4, 6,7, 4,4, 2,3, 2,0, und 0,6 kb). Die Spuren 2 bis 10 enthalten genomische DNA von Mensch, Rhesusaffe, Spraque Dawley Ratte, BALB/c Maus, Hund, Rind, Kaninchen, Huhn und Hefe (*Saccharomyces cerevisiae*).

B: Zu Fig.1 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse, hybridisiert mit einer radioaktiv-markierten etwa 720 bp langen hTC-cDNA Sonde.

Fig. 2: Restriktionsanalyse der rekombinanten λ-DNA des Phagenklons P12, der mit einer Sonde aus dem 5′-Bereich der hTC-cDNA hybridisiert.

Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,4 %igen Agarosegels. Die Spuren 1 und 2 enthalten Eco RI/Hind III geschnittene λ-DNA bzw eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 3 - 7 enhalten 250 ng mit Bam HI (Spur 3), Eco RI (Spur 4), Sal I (Spur 5), Xho I (Spur 6) und Sac I (Spur 7) geschnittene DNA des rekombinanten Phagens. Die Pfeile kennzeichnen die zwei λ-Arme des Vektors EMBL3 Sp6/T7.

Fig. 3: Restriktionsanalyse und Southern Blot-Analyse der rekombinanten λ-DNA des Phagenklons, der mit einer Sonde aus dem 5′-Bereich der hTC-cDNA hybridisiert.

A: Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,8%igen Agarosegels. Die Spuren 1 und 15 enthalten eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 2 bis 14 enthalten 250 ng geschnittene λ-DNA vom rekombinanten Phagenklon. Als Enzyme wurden eingesetzt: Spur 2: Sac I, Spur 3: Xho I, Spur 4: Xho I, Xba I, Spur 5: Sac I, Xho I, Spur 6: Sal I, Xho I, Xba I, Spur 7: Sac I, Xho I, Xba I, Spur 8: Sac I, Sal I, Xba I, Spur 9: Sac I, Sal I, BamH I, Spur 10: Sac I, Sal I, Xho I, Spur 11: Not I, Spur 12: Sma I, Spur 13: leer, Spur 14: nicht verdaut.

B: Zu Fig.3 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse. Als Sonde für die Hybridisierung wurde ein etwa 420 bp langes 5′-hTC-cDNA Fragment eingesetzt.

Fig. 4: Partielle DNA-Sequenz der 5′-flankierenden Region und des Promotors vom Gen der humanen katalytischen Telomerase-Untereinheit. Das ATG-Startcodon ist in der Sequenz fett hervorgehoben. Die dargestellte Sequenz entspricht SEQ ID NO 1.

Fig. 5: Identifizierung des Transkriptionsstarts durch Primer Extension-Analyse.

Die Abbildung zeigt ein Autoradiogramm eines denaturierenden Polyacrylamidgels, welches zur Darstellung einer Primer Extension-Analyse gewählt wurde. Als Primer wurde ein Oligonukleotid mit der Sequenz 5′GTTAAGTTGTAGCTTACACTGGTTCTC 3′ benutzt. In der Spur 1 wurde die Primer Extension Reaktion aufgetragen. Die Spuren G, A, T, C, stellen die Sequenzreaktionen mit dem gleichen Primer und den entsprechenden Dideoxynukleotiden dar. Der fette Pfeil kennzeichnet den Haupt-Transkriptionsstart, die dünnen Pfeile weisen auf drei Neben-Transkriptionsstartpunkte hin.

Fig. 6: cDNA Sequenz der humanen katalytischen Telomerase-Untereinheit (hTC; vgl. unsere anhängige Anmeldung PCT/EP/98/03468). Die dargestellte Sequenz entspricht SEQ ID NO 2.

Fig. 7: Strukturelle Organisation und Restriktionsmappe des humanen hTC-Gens und dessen 5′- und 3′-flankierende Region.

Exons sind als durchnummerierte schwarz ausgefüllte Rechtecke und Introns als nicht ausgefüllte Bereiche hervorgehoben. Nichttranslatierte Sequenzabschnitte in den Exons sind schraffiert. Die Translation startet in Exon 1 und endet in Exon 16. Restriktionsenzymschnittstellen sind wie folgt gekennzeichnet: S, SacI; X, XhoI. Die relative Anordnung der fünf Phagenklone (P2, P3, P5, P12, P17) und des Produktes aus dem „Genomic walking“ sind durch dünne Linien hervorgehoben. Wie durch

die Punkte gekennzeichnet, ist die Sequenz von Intron 16 nur teilweise entschlüsselt.

Fig. 8: HTL Splicevarianten.

A: Schematische Struktur der hTC mRNA Splicevarianten. Die vollständige hTC mRNA ist als grau unterlegtes Rechteck im oberen Bereich der Abb. dargestellt. Die 16 Exons sind entsprechend ihrer Größe dargestellt. Der Translationsstart (ATG) und das Stop-Codon, sowie das Telomerase-spezifische T-Motiv und die sieben RT-Motive sind hervorgehoben. Die hTC-Varianten sind in Deletions- und Insertionsvarianten unterteilt. In den Deletionen sind die fehlenden Exonsequenzen markiert. Die Insertionen sind durch zusätzliche weiße Rechtecke hervorgehoben. Größe und Herkunft der insertierten Sequenzen sind angegeben. Neu entstandene Stop-Codons sind markiert. Die Größe der Insertion von Variante INS2 ist unbekannt.

B: Exon Intron Übergänge der hTC-Splice-Varianten. Nichtgesplicte 5′- und 3′-flankierte Sequenzen sind als weiße Rechtecke hervorgehoben. Die Herkunft der Exon und Intron Sequenzen ist angegeben. Intron und Exon Sequenzen sind in Kleinbuchstaben, bzw. in Großbuchstaben dargestellt. Die Donor und Akzeptor Sequenzen der Splicestellen sind als graue Rechtecke unterlegt und deren Exon Intron Herkunft ist ebenfalls angegeben.

Fig. 9: Identifizierung des Transkriptionsstarts durch RT-PCR Analyse.
Die RT-PCR wurde mit cDNA-Bibliothek aus HL 60 Zellen und genomischer DNA als Positivkontrolle durchgeführt. Ein gemeinsamer 3′-Primer hybridisiert an eine Sequenzregion aus Exon 1. Die Position der verschiedenen 5′ Primer in der kodierenden Region oder der 5′-flankierenden Region ist angegeben. In der Negativkontrolle wurde keine

Template-DNA in der PCR-Reaktion zugegeben. M: DNA-Größenmarker.

Fig. 10: Nukleotidsequenz und Strukturmerkmale des hTC-Promotors.
11273 bp der 5'-flankierenden hTC Gensequenz, beginnend mit dem Translationsstartcodon ATG (+1) sind dargestellt. Die putative Region des Translationsstarts ist unterstrichen. Mögliche regulatorische Sequenzabschnitte innerhalb der 4000 bp stromaufwärts des Translationsstarts sind umrandet. Die dargestellte Sequenz entspricht SEQ ID NO 3.

Fig. 11: Aktivität des hTC-Promotors in HEK-293 Zellen.
Im oberen Bereich der Abbildung sind die ersten 5000 bp der 5'-flankierenden hTC Genregion schematisch dargestellt. Das ATG-Startcodon ist hervorgehoben. CpG reiche Inseln sind durch graue Rechtecke markiert. Auf der linken Abbildungsseite sind die Größen der hTC Promotor-Luziferase Konstukte dargestellt. Das pomotorlose pGL2-Basic Konstrukt und das SV40 Promotorkonstrukt pGL2-Pro wurden in jeder Transfektion als Kontrollen eingesetzt. Auf der rechten Abbildungsseite sind ist die relative Luziferaseaktivität der verschiedenen Promotorkonstukte in HEK-Zellen als durchgehende Balken gezeigt. Die Standardabweichung ist angegeben. Die Zahlenwerte repräsentieren den Durchschnitt von zwei unabhängigen Experimenten, die in Duplikaten durchgeführt wurden.

Tab. 1: Exon Intron Übergänge des hTC-Gens
Aufgelistet sind die Nukleotidsequenzen an den 3'- und 5' Spliceübergängen des hTC-Gens. Die Konsensussequenzen für Donor und Akzeptorsequenzen (AG und GT) sind durch graue Rechtecke unterlegt. Intronsequenzen (Kleinbuchstaben) und Exonsequenzen (Großbuchstaben), die die Spliceakzeptor- und Donorstellen flankieren sind gezeigt. Die Größe der Exons und Introns ist in bp angegeben.

Tab. 2: Potentielle Bindungsstellen für DNA-bindende Faktoren in der Nukleotidsequenz von Intron 2

Die Suche nach möglichen DNA-bindenden Faktoren (z.B. Transkriptionsfaktoren) wurde mit dem „Find Pattern"-Algorithmuses aus dem „GCG Sequenz Analysis" Programmpacket der „Genetics Computer Group" (Madison, USA) durchgeführt. Aufgelistet sind die Abkürzungen der identifizierten DNA-bindenden Faktoren und deren Lokalisation in Intron 2.

Tab. 1

3' Acceptor Sequence				5' Donor Sequence			
Intron	Exon	Exon No.	bp	Exon	Intron	Intron No.	bp
5' flankierende Region	GTTTCAGGCAGCGCTGCGT	1	281	CGCCCCCTCCTTCCGCCAG	gtgggcctccccggggtcg	1	104
cagggcgcttcccccgcag	GTGTCCTGCCTGAAGGAGC	2	1354	TGGCTGCGCAGGAGCCCAG	gtgaggaggtggtggccgt	2	8616
catgtccttctcgtttaag	GGGTTGGCTGTGTTCCGGC	3	196	TGCAAAGCATTGGAATCAG	gtactgtatccccacgcca	3	2089
gaggggctctctattgcag	ACAGCACTTGAAGAGGGTG	4	181	GTTCCGCAGAGAAAAGAGG	gtggctgtgctttggttta	4	687
cccatgctgtccccgccag	GCCGAGCGTCTCACCTCGA	5	180	TGAGCTGTACTTTGTCAAG	gtgggtgccggggaccccc	5	494
ctcgcctccactcacacag	GTGGATGTGACGGGCGCGT	6	156	CAAGGCCTTCAAGAGCCAC	gtaaggttcacgtgtgata	6	>4660
ccctctcctctgccggcag	GTCTCTACCTTGACAGACC	7	96	TGCCGTCGTCATCGAGCAG	gtctgggcactgccctgca	7	980
ctcccgtctgctttcgcag	AGCTCCTCCCTGAATGAGG	8	86	CCGTGCGCATCAGGGGCAA	gtgagtcaggtggccaggt	8	2485
ctgtgtcttcccgccccag	GTCCTACGTCCAGTGCCAG	9	114	CGGGGATTCGGCGGGACGG	gtgaggcctcctcttcccc	9	1984
gtattttcccttatttcag	GCTGCTCCTGCGTTTGGTG	10	72	ACGCGAAAACCTTCCTCAG	gtgaggcccgtgccgtgtg	10	1871
cattgcccctctgccttag	GACCCTGGTCCGAGGTGTC	11	189	TGCAGAGCGACTACTCCAG	gtgagcgcacctggccgga	11	3803
attccccccctgtgtctcag	CTATGCCCGGACCTCCATC	12	127	CCTGTTTCTGGATTTGCAG	gtgagcaggctgatggtca	12	880
tctttcttggcgactctag	GTGAACAGCCTCCAGACGG	13	62	TCCTGCTGCAGGCGTACAG	gtgagccgccaccaagggg	13	3187
ctgtccgccatcctctcag	GTTTCACGCATGTGTGCTG	14	125	CTGAAAGCCAAGAACGCAG	gtatgtgcaggtgcctggc	14	781
agcctctgttttcccccag	GGATGTCGCTGGGGGCCAA	15	138	CTGGGGTCACTCAGGACAG	gcaagtgtgggtggaggcc	15	536
tctgattttggccccgcag	CCCAGACGCAGCTGAGTCG	16	664	TTTTTCAGTTTTGAAAAAA	3' flankierende Region		

Tab. 2

Faktoren	**Lokalisation in Intron 2**
C/EBP	2925
CRE.2	2749
Sp1	2378, 4094, 4526, 4787, 4835, 4995
AP-2 CS3	5099
AP-2 CS4	2213, 3699, 4667, 5878, 5938, 6059, 6180, 6496
AP-2 CS5	5350, 5798, 5880, 5940, 6061, 6182, 6375, 6498
PEA3	934, 2505
P53	2125
GR uteroglobin	848, 1487, 2956
PR uteroglobin	3331
Zeste-white	1577, 1619, 1703, 1745, 1787, 1829, 1871, 1913, 1955, 1997, 2039, 2081, 3518, 3709, 4765, 5014, 5055
GRE	846
MyoD-MCK right site/rev	447, 509, 558, 1370, 1595, 1900, 2028, 2099, 4557
MyoD-MCK left site	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
Ets-1 CS	6408
AP1	3784, 4406
CREB	2801
GATA-1	839, 1390, 3154
c-Myc	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
CACCC site	991
CCAAT site	1224
CCAC box	992
CAAT site	463, 2395
Rb site	992, 4663
TATA	3650
CDEI	106, 1564, 1606, 1690, 1732, 1816, 1900, 1984

Beispiele

Das menschliche Gen für die katalytische Telomerase Untereinheit (ghTC), sowie die 5′ und 3′ liegenden Bereiche dieses Gens wurden kloniert, der Startpunkt der Transkription bestimmt, potentielle Bindungsstellen für DNA-bindende Proteine identifiziert, sowie aktive Promotorfragmente aufgezeigt. Die Sequenz der hTC-cDNA (Fig. 6) ist bereits in unserer ebenfalls anhängigen Anmeldung PCT/EP/98/03468 beschrieben. Wenn nicht gesondert erwähnt, beziehen sich sämtliche Angaben zur cDNA-Position auf diese Sequenz.

Beispiel 1

Durch eine genomische Southern Blot-Analyse wurde bestimmt, ob ghTC im menschlichen Genom ein Einzelgen darstellt oder mehrere Loci für das hTC-Gen bzw. eventuell auch ghTC-Pseudogene existieren.

Hierzu wurde ein kommerziell erhältlicher Zoo-Blot der Firma Clontech einer Southern Blot-Analyse unterzogen. Dieser Blot enthält 4 µg Eco RI geschnittene genomische DNA von neun verschiedenen Spezies (Mensch, Affe, Ratte, Maus, Hund, Rind, Kaninchen, Huhn und Hefe). Mit Ausnahme von Hefe, Huhn und Mensch wurde die DNA aus Nierengewebe isoliert. Die humane genomische DNA wurde aus Plazenta isoliert und die genomische DNA aus Huhn wurde aus Lebergewebe aufgereinigt. Im Autoradiogramm in Fig. 1 wurde als radioaktiv-markierte Sonde ein etwa 720 bp langes hTC-cDNA Fragment, isoliert aus der hTC cDNA, Variante Del2 (Position 1685 bis 2349 plus 2531 bis 2590 der Fig. 6 [Deletion 2; vergl. Beispiel 5 der Fig. 8]), eingesetzt. Die experimentellen Bedingungen für die Hybridisierung und die Waschschritte des Blots erfolgten in Anlehnung an Ausubel *et al.* (1987).

Im Fall der humanen DNA erkennt die Sonde zwei spezifische DNA-Fragmente. Das kleinere, etwa 1,5 bis 1,8 kb lange Eco RI-Fragment geht wahrscheinlich auf zwei

Eco RI-Schnittstellen in einem Intron der ghTC-DNA zurück. Aufgrund dieses Ergebnisses ist davon auszugehen, daß nur ein singuläres ghTC-Gen im menschlichen Genom vorliegt.

Beispiel 2

Zur Isolierung der 5′ flankierenden hTC-Gensequenz wurden ca 1,5 x 10^6 Phagen einer humanen genomischen Plazenta-Genbibliothek (EMBL 3 SP6/T7 der Firma Clontech, Bestellnummer HL1067j) auf Nitrozellulosefilter (0,45 µm; Fa. Schleicher und Schuell) nach Angaben des Herstellers mit einem radioaktiv markierten, etwa 500 bp langen 5′-hTC-cDNA Fragment (Position 839 bis 1345 der Fig. 6) hybridisiert. Die Nitrozellulosefilter wurden zunächst in 2 x SSC (0,3 M NaCl; 0,5 M Tris-HCl, pH 8,0) und anschließend in einer Prähybridisierungslösung (50 % Formamid; 5 x SSPE, pH 7,4; 5 x Denhards-Lösung; 0,25 % SDS; 100 µg/ml Heringsperma-DNA) zwei Stunden bei 42°C inkubiert. Für die Hybrididsierung über Nacht wurde die Prähybridisierungslösung mit 1,5 x 10^6 cpm/ml Lösung denaturierter, radioaktiv markierter Probe ergänzt. Unspezifisch gebundene, radioaktive DNA wurde unter stringenten Bedingungen, d.h. durch drei fünfminütige Waschschritte mit 2 x SSC; 0,1 % SDS bei 55 bis 65 °C entfernt. Die Auswertung erfolgte durch Autoradiographie der Filter.

Die in dieser Primäruntersuchung identifizierten Phagenklone wurden aufgereinigt Ausubel *et al.* (1987). In weitergehenden Analysen stellte sich ein Phagenklon P12 als potentiell positiv heraus. Eine λ-DNA Präparation dieses Phagens Ausubel *et al.* (1987) und der nachfolgende Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, zeigte, daß dieser Phagenklon ein ca. 15 kb Insert im Vektor enthält (Fig. 2).

Zur Isolierung der vollständigen hTC-Gensequenz wurden in unabhängigen Experimenten jeweils 1 bis 1,5 x 10^6 Phagen mit jeweils verschiedenen radioaktiv markierten Sonden wie oben beschrieben durchmustert.

Die in diesen Primäruntersuchungen identifizierten, für die entsprechenden Sonden positiven Phagenklone wurden aufgereinigt. Der Phagenklon P17 wurde mit einem etwa 250 bp langen hTC-cDNA Fragment (Position 1787 bis 2040 der Fig. 6) gefunden. Der Phagenklon P2 wurde mit einem etwa 740 bp langen hTC-cDNA Fragment (Position 1685 bis 2349 plus 2531 bis 2607 der Fig. 6 [Deletion 2; vergl. Beispiel 5]) identifiziert. Die Phagenklone P3 und P5 wurden mit einem 420 bp langen 3′ hTC-cDNA Fragment (Position 3047 bis 3470 der Fig. 6) gefunden. Nach λ-DNA Präparation dieser Phagen und nachfolgendem Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, wurden die Inserts in Plasmide umkloniert (Beispiel 4).

Beispiel 3

Um zu untersuchen, ob auch das 5′-Ende der hTC-cDNA im Insert des rekombinanten Phagenklons P12 vorliegt, wurde λ-DNA dieses Klons in einer Southern Blot Analyse mit einem radioaktiv markierten etwa 440 bp langen hTC-cDNA Fragment (Position 1 bis 440 der Fig. 6) aus dem extremen 5′-Bereich hybridisiert (Fig. 3).

Da die isolierte λ-DNA des positiven Klons auch mit dem extremen 5′-Ende der hTC-cDNA hybridisiert, enthält dieser Phage wahrscheinlich auch den das ATG-Startcodon flankierenden 5′-Sequenzbereich.

Beispiel 4

Um das gesamte 15 kb lange Insert des positiven Phagenklons P12 in Teilfragmenten umzuklonieren und anschließend zu sequenzieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einem das gesamte Insert aus EMBL3 Sp6/T7 freisetzen (vgl. Beispiel 2) und zusätzlich im Insert schneiden.

Insgesamt wurden ein etwa 8,3 und ein etwa 6,5 kb langes Xho I-Subfragment sowie ein etwa 8,5, ein etwa 3,5 und ein etwa 3 kb langes Sac I-Teilfragment in den Vektor pBluescript KS(+) (Fa. Stratagene) umkloniert. Durch Sequenzanalyse dieser Fragmente wurde die Nukleotidsequenz von 5123 bp 5′-flankierenden des ghTC-Genbereichs, ausgehend vom ATG-Startcodon bestimmt (Fig. 4; entsprechend SEQ ID NO 1). In der Fig. 4 sind die ersten (ausgehend vom ATG-Startcodon) 5123 bp dargestellt. In der Fig. 10 (entsprechend SEQ ID NO 3) die gesamte klonierte 5′ Sequenz.

Um das gesamte ca. 14,6 kb große Insert des Phagenklons P17 in Teilfragmenten umzuklonieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einen das gesamte Insert aus EMLB3 Sp6/T7 freisetzen und zusätzlich einige Male im Insert schneiden. Durch Kombinationsverdau mit den Enzymen XhoI und BamHI wurden ein 7,1 kb, ein 4,2 kb und ein 1,5 kb großes XhoI-BamHI-Fragment sowie ein 1,8 kb großes BamHI-Fragment subkloniert. Der Kombinations-Restriktionsverdau mit den Enzymen XhoI und XbaI führte zur Klonierung von einem 6,5 kb großen XhoI-XbaI-Fragment, einem 6,5 kb und einem 1,5 kb großem XhoI-Fragment.

Die Umklonierung des ca. 17,9 kb großem Inserts des Phagenklons P2 in Subfragmente erfolgte durch Verdau mit dem Restriktionsenzym XhoI. Insgesamt wurde ein 7,5 kb, ein 6,4 kb sowie ein 1,6 kb langes XhoI-Subfragment kloniert. Durch Verdau mit dem Restriktionsenzym SacI wurde zusätzlich ein 4,8 kb, ein 3 kb, ein 2 kb sowie ein 1,8 kb großes SacI-Fragment subkloniert.

Das ca. 13,5 kb große Insert des Phagenklons P3 wurde durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI subkloniert. Dabei wurden ein 3,2 kb, ein 2 kb, ein 0,9 kb, ein 0,8 kb, ein 0,65 kb und ein 0,5 kb langes SacI-Subfragment sowie ein 6,5 kb und ein 4,3 kb langes XhoI-Subfragment erhalten.

Die Subklonierung des ca. 13,2 kb großen Inserts des Phagenklons P5 erfolgte durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI. Insgesamt wurden SacI-Fragmente von 6,5 kb, 3,3 kb, 3,2 kb, 0,8 kb und 0,3 kb Größe sowie XhoI-Fragmente von 7 kb und 3,2 kb Größe subkloniert.

Zur Klonierung des 3' von Phagenklon P17 und 5' von Phagenklon P2 gelegenen hTC-genomischen Sequenzbereichs wurden 3 Genomic Walkings mit Hilfe des GenomeWalker™ Kits der Firma Clontech (Katalognummer K1803-1) und verschiedenen Primerkombinationen durchgeführt. In einem Endvolumen von 50 µl wurde 1 µl humaner GenomeWalker Library HDL (Fa. Clontech) mit 10 pmol dNTP-Mix versetzt und in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase Mix (Fa. Clontech) eine PCR-Reaktion durchgeführt. Als Primer wurden 10 pmol eines internen genspezifischen Primers sowie 10 pmol des Adaptor Primers AP1 (5'-GTAATACGACTCACTATAGGGC-3'; Fa. Clontech) zugefügt. Die PCR wurde als Touchdown-PCR in 3 Schritten durchgeführt. Zunächst wurde über 7 Zyklen für 20 sec bei 94°C denaturiert und anschließend für 4 min bei 72°C die Primer angelagert und die DNA-Kette verlängert. Es folgten 37 Zyklen bei denen für 20 sec die DNA bei 94°C denaturiert wurde, die anschließende Primerverlängerung aber für 4 min bei 67°C erfolgte. Abschließend folgte eine Kettenverlängerung für 4 min bei 67°C. Im Anschluß an diese erste PCR wurde das PCR-Produkt 1:50 verdünnt. Ein µl dieser Verdünnung wurde in einer zweiten „nested" PCR zusammen mit 10 pmol dNTP-Mix in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase-Mix sowie 10 pmol eines „nested" genspezifischen Primers und 10 pmol des „nested" Marathon Adaptor Primers AP2 (5'-ACTATAGGGCACGCGTGGT-3'; Fa. Clontech) eingesetzt. Die PCR-Bedingungen entsprachen den in der ersten PCR gewählten Parametern. Als einzige Ausnahme wurden im ersten PCR-Schritt statt 7 Zyklen nur 5 Zyklen gewählt und im zweiten PCR-Schritt statt 37 Zyklen nur 24 Zyklen durchlaufen. Produkte dieser Nested-GenomicWalking-PCR wurden in den TA-Cloning Vektor pCRII der Fa. InVitrogen kloniert.

Im ersten Genomic Walking wurde der genspezifische Primer C3K2-GSP1 (5'-GACGTGGCTCTTGAAGGCCTTG-3') sowie der „nested" genspezifische Primer C3K2-GSP2 (5'-GCCTTCTGGACCACGGCATACC-3') zusammen mit der HDL-Library 4 eingesetzt und ein 1639 bp langes PCR-Fragment erhalten. Im zweiten Genomic Walking wurde mit dem genspezifischen Primer C3F2 (5'-CGTAGTTGAGCACGCTGAACAGTG-3') und dem „nested" genspezifischen Primer C3F (5'-CCTTCACCCTCGAGGTGAGACGCT-3') aus der HDL-Library 4 ein PCR-Fragment von 685 bp Länge amplifiziert. Der dritte Genomic Walking Ansatz führte unter Einsatz des genspezifischen Primers DEL5-GSP1 (5'-GGTGGATGTGACGGGCGCGTACG-3') und des „nested" genspezifischen Primers C5K-GSP1 (5'-GGTATGCCGTGGTCCAGAAGGC-3') zur Klonierung eines 924 bp PCR-Fragments aus der HDL-Library 1. Insgesamt wurden durch dieses Genomic Walking-Verfahren 2100 bp der 3' von Phagenklon P17 gelegenen genomischen hTC-Region identifiziert (s. Fig. 7).

Die subklonierten Fragmente sowie die Genomic Walking-Produkte wurden einzelsträngig sequenziert. Unter Verwendung der Lasergene Biocomputing Software (DNASTAR Inc. Madison, Wisconsin, USA) wurden überlappende Bereiche identifiziert und Contigs gebildet. Insgesamt wurden aus den gesammelten Sequenzen der Phagenklone P12, P17, P2, P3 und P5 sowie den Sequenzdaten aus dem Genomic Walking 2 große Contigs zusammengestellt. Contig 1 besteht aus Sequenzdaten von Phagenklon P12, P17 und den Sequenzdaten aus dem Genomic Walking. Contig 2 wurde aus den Sequenzen von Phagenklon P2, P3 und P5 zusammengesetzt. Überlappende Phagenklonbereiche sind in Fig. 7 schematisch dargestellt. Die Sequenzdaten der 2 Contigs sind nachfolgend dargestellt. Das ATG Startcodon in Contig 1 ist unterstrichen. Das TGA Stopcodon ist in Contig 2 unterstrichen.

Contig1:

```
ACTTGAGCCC AAGAGTTCAA GGCTACGGTG AGCCATGATT GCAACACCAC ACGCCAGCCT TGGTGACAGA 70
ATGAGACCCT GTCTCAAAAA AAAAAAAAAA AATTGAAATA ATATAAAGCA TCTTCTCTGG CCACAGTGGA 140
ACAAAACCAG AAATCAACAA CAAGAGGAAT TTTGAAAACT ATACAAACAC ATGAAAATTA AACAATATAC 210
TTCTGAATGA CCAGTGAGTC AATGAAGAAA TTAAAAAGGA AATTGAAAAA TTTATTTAAG CAAATGATAA 280
CGGAAACATA ACCTCTCAAA ACCCACGGTA TACAGCAAAA GCAGTGCTAA GAAGGAAGTT TATAGCTATA 350
AGCAGCTACA TCAAAAAAGT AGAAAAGCCA GGCGCAGTGG CTCATGCCTG TAATCCCAGC ACTTTGGGAG 420
GCCAAGGCGG GCAGATCGCC TGAGGTCAGG AGTTCGAGAC CAGCCTGACC AACACAGAGA AACCTTGTCG 490
CTACTAAAAA TACAAAATTA GCTGGGCATG GTGGCACATG CCTGTAATCC CAGCTACTCG GGAGGCTGAG 560
GCAGGATAAC CGCTTGAACC CAGGAGGTGG AGGTTGCGGT GAGCCGGGAT TGCGCCATTG GACTCCAGCC 630
TGGGTAACAA GAGTGAAACC CTGTCTCAAG AAAAAAAAAA AAGTAGAAAA ACTTAAAAAT ACAACCTAAT 700
GATGCACCTT AAAGAACTAG AAAAGCAAGA GCAAACTAAA CCTAAAATTG GTAAAAGAAA AGAAATAATA 770
AAGATCAGAG CAGAAATAAA TGAAACTGAA AGATAACAAT ACAAAAGATC AACAAAATTA AAAGTTGGTT 840
TTTTGAAAAG ATAAACAAAA TTGACAAACC TTTGCCCAGA CTAAGAAAAA AGGAAAGAAG ACCTAAATAA 910
ATAAAGTCAG AGATGAAAAA AGAGACATTA CAACTGATAC CACAGAAATT CAAAGGATCA CTAGAGGCTA 980
CTATGAGCAA CTGTACACTA ATAAATTGAA AAACCTAGAA AAAATAGATA AATTCCTAGA TGCATACAAC 1050
CTACCAAGAT TGAACCATGA AGAAATCCAA AGCCCAAACA GACCAATAAC AATAATGGGA TTAAAGCCAT 1120
AATAAAAAGT CTCCTAGCAA AGAGAAGCCC AGGACCCAAT GGCTTCCCTG CTGGATTTTA CCAATCATTT 1190
AAAGAAGAAT GAATTCCAAT CCTACTCAAA CTATTCTGAA AAATAGAGGA AAGAATACTT CCAAACTCAT 1260
TCTACATGGC CAGTATTACC CTGATTCCAA AACCAGACAA AAACACATCA AAAACAAACA AACAAAAAAA 1330
CAGAAAGAAA GAAAACTACA GGCCAATATC CCTGATGAAT ACTGATACAA AAATCCTCAA CAAAACACTA 1400
GCAAACCAAA TTAAACAACA CCTTCGAAAG ATCATTCATT GTGATCAAGT GGGATTTATT CCAGGGATGG 1470
AAGGATGGTT CAACATATGC AAATCAATCA ATGTGATACA TCATCCCAAC AAAATGAAGT ACAAAAACTA 1540
TATGATTATT TCACTTTATG CAGAAAAAGC ATTTGATAAA ATTCTGCACC CTTCATGATA AAAACCCTCA 1610
AAAAACCAGG TATACAAGAA ACATACAGGC CAGGCACAGT GGCTCACACC TGCGATCCCA GCACTCTGGG 1680
AGGCCAAGGT GGGATGATTG CTTGGGCCCA GGAGTTTGAG ACTAGCCTGG GCAACAAAAT GAGACCTGGT 1750
CTACAAAAAA CTTTTTTAAA AAATTAGCCA GGCATGATGG CATATGCCTG TAGTCCCAGC TAGTCTGGAG 1820
GCTGAGGTGG GAGAATCACT TAAGCCTAGG AGGTCGAGGC TGCAGTGAGC CATGAACATG TCACTGTACT 1890
CCAGCCTAGA CAACAGAACA AGACCCCACT GAATAAGAAG AAGGAGAAGG AGAAGGGAGA AGGGAGGGAG 1960
AAGGGAGGAG GAGGAGAAGG AGGAGGTGGA GGAGAAGTGG AAGGGGAAGG GGAAGGGAAA GAGGAAGAAG 2030
AAGAAACATA TTTCAACATA ATAAAAGCCC TATATGACAG ACCGAGGTAG TATTATGAGG AAAAACTGAA 2100
AGCCTTTCCT CTAAGATCTG GAAAATGACA AGGGCCCACT TTCACCACTG TGATTCAACA TAGTACTAGA 2170
AGTCCTAGCT AGAGCAATCA GATAAGAGAA AGAAATAAAA GGCATCCAAA CTGGAAAGGA AGAAGTCAAA 2240
TTATCCTGTT TGCAGATGAT ATGATCTTAT ATCTGGAAAA GACTTAAGAC ACCACTAAAA AACTATTAGA 2310
GCTGAAATTT GGTACAGCAG GATACAAAAT CAATGTACAA AAATCAGTAG TATTTCTATA TTCCAACAGC 2380
AAACAATCTG AAAAAGAAAC CAAAAAAGCA GCTACAAATA AAATTAAACA GCTAGGAATT AACCAAAGAA 2450
GTGAAAGATC TCTACAATGA AAACTATAAA ATGTTGATAA AAGAAATTGA AGAGGGCACA AAAAAAGAAA 2520
AGATATTCCA TGTTCATAGA TTGGAAGAAT AAATACTGTT AAAATGTCCA TACTACCCAA AGCAATTTAC 2590
AAATTCAATG CAATCCCTAT TAAAATACTA ATGACGTTCT TCACAGAAAT AGAAGAAACA ATTCTAAGAT 2660
TTGTACAGAA CCACAAAAGA CCCAGAATAG CCAAAGCTAT CCTGACCAAA AAGAACAAAA CTGGAAGCAT 2730
CACATTACCT GACTTCAAAT TATACTACAA AGCTATAGTA ACCCAAACTA CATGGTACTG GCATAAAAAC 2800
AGATGAGACA TGGACCAGAG GAACAGAATA GAGAATCCAG AAACAAATCC ATGCATCTAC AGTGAACTCA 2870
TTTTTGACAA AGGTGCCAAG AACATACTTT GGGGAAAAGA TAATCTCTTC AATAAATGGT GCTGGAGGAA 2940
CTGGATATCC ATATGCAAAA TAACAATACT AGAACTCTGT CTCTCACCAT ATACAAAAGC AAATCAAAAT 3010
GGATGAAAGG CTTAAATCTA AAACCTCAAA CTTTGCAACT ACTAAAAGAA AACACCGGAG AAACTCTCCA 3080
GGACATTGGA GTGGGCAAAG ACTTCTTGAG TAATTCCCTG CAGGCACAGG CAACCAAAGC AAAAACAGAC 3150
AAATGGGATC ATATCAAGTT AAAAAGCTTC TGCCCAGCAA AGGAAACAAT CAACAAAGAG AAGAGACAAC 3220
CCACAGAATG GGAGAATATA TTTGCAAACT ATTCATCTAA CAAGGAATTA ATAACCAGTA TATATAAGGA 3290
GCTCAAACTA CTCTATAAGA AAAACACCTA ATAAGCTGAT TTTCAAAAAT AAGCAAAAGA TCTGGGTAGA 3360
CATTTCTCAA AATAAGTCAT ACAAATGGCA AACAGGCATC TGAAAATGTG CTCAACACCA CTGATCATCA 3430
GAGAAATGCA AATCAAAACT ACTATGAGAG ATCATCTCAT CCCAGTTAAA ATGGCTTTTA TTCAAAAGAC 3500
AGGCAATAAC AAATGCCAGT GAGGATGTGG ATAAAAGGAA ACCCTTGGAC ACTGTTGGTG GGAATGGAAA 3570
TTGCTACCAC TATGGAGAAC AGTTTGAAAG TTCCTCAAAA AACTAAAAAT AAAGCTACCA TACAGCAATC 3640
CCATTGCTAG GTATATACTC CAAAAAAGGG AATCAGTGTA TCAACAAGCT ATCTCCACTC CCACATTTAC 3710
TGCAGCACTG TTCATAGCAG CCAAGGTTTG GAAGCAACCT CAGTGTCCAT CAACAGACGA ATGGAAAAAG 3780
AAAATGTGGT GCACATACAC AATGGAGTAC TACGCAGCCA TAAAAAAGAA TGAGATCCTG TCAGTTGCAA 3850
CAGCATGGGG GGCACTGGTC AGTATGTTAA GTGAAATAAG CCAGGCACAG AAAGACAAAC TTTTCATGTT 3920
CTCCCTTACT TGTGGGAGCA AAAATTAAAA CAATTGACAT AGAAATAGAG GAGAATGGTG GTTCTAGAGG 3990
GGTGGGGGAC AGGGTGACTA GAGTCAACAA TAATTTATTG TATGTTTTAA AATAACTAAA AGAGTATAAT 4060
TGGGTTGTTT GTAACACAAA GAAAGGATAA ATGCTTGAAG GTGACAGATA CCCCATTTAC CCTGATGTGA 4130
TTATTACACA TTGTATGCCT GTATCAAAAT ATCTCATGTA TGCTATAGAT ATAAACCCTA CTATATTAAA 4200
AATTAAAATT TTAATGGCCA GGCACGGTGG CTCATGTCCG TAATCCCAGC ACTTTGGGAG GCCGAGGCGG 4270
GTGGATCACC TGAGGTCAGG AGTTTGAAAC CAGTCTGGCC ACCATGATGA AACCCTGTCT CTACTAAAGA 4340
TACAAAAATT AGCCAGGCGT GGTGGCACAT ACCTGTAGTC CCAACTACTC AGGAGGCTGA GACAGGAGAA 4410
TTGCTTGAAC CTGGGAGGCG GAGGTTGCAG TGAGCCGAGA TCATGCCACT GCACTGCAGC CTGGGTGACA 4480
GAGCAAGACT CCATCTCAAA ACAAAAACAA AAAAAAGAAG ATTAAAATTG TAATTTTTAT GTACCGTATA 4550
AATATATACT CTACTATATT AGAAGTTAAA AATTAAAACA ATTATAAAAG GTAATTAACC ACTTAATCTA 4620
AAATAAGAAC AATGTATGTG GGGTTTCTAG CTTCTGAAGA AGTAAAAGTT ATGGCCACGA TGGCAGAAAT 4690
GTGAGGAGGG AACAGTGGAA GTTACTGTTG TTAGACGCTC ATACTCTCTG TAAGTGACTT AATTTTAACC 4760
AAAGACAGGC TGGGAGAAGT TAAAGAGGCA TTCTATAAGC CCTAAAACAA CTGCTAATAA TGGTGAAAGG 4830
TAATCTCTAT TAATTACCAA TAATTACAGA TATCTCTAAA ATCGAGCTGC AGAATTGGCA CGTCTGATCA 4900
CACCGTCCTC TCATTCACGG TGCTTTTTTT CTTGTGTGCT TGGAGATTTT CGATTGTGTG TTCGTGTTTG 4970
GTTAAACTTA ATCTGTATGA ATCCTGAAAC GAAAAATGGT GGTGATTTCC TCCAGAAGAA TTAGAGTACC 5040
TGGCAGGAAG CAGGTGGCTC TGTGGACCTG AGCCACTTCA ATCTTCAAGG GTCTCTGGCC AAGACCCAGG 5110
```

TGCAAGGCAG AGGCCTGATG ACCCGAGGAC AGGAAAGCTC GGATGGGAAG GGGCGATGAG AAGCCTGCCT 5180
CGTTGGTGAG CAGCGCATGA AGTGCCCTTA TTTACGCTTT GCAAAGATTG CTCTGGATAC CATCTGGAAA 5250
AGGCGGCCAG CGGGAATGCA AGGAGTCAGA AGCCTCCTGC TCAAACCCAG GCCAGCAGCT ATGGCGCCCA 5320
CCCGGGCGTG TGCCAGAGGG AGAGGAGTCA AGGCACCTCG AAGTATGGCT TAAATCTTTT TTTCACCTGA 5390
AGCAGTGACC AAGGTGTATT CTGAGGGAAG CTTGAGTTAG GTGCCTTCTT TAAAACAGAA AGTCATGGAA 5460
GCACCCTTCT CAAGGGAAAA CCAGACGCCC GCTCTGCGGT CATTTACCTC TTTCCTCTCT CCCTCTCTTG 5530
CCCTCGCGGT TTCTGATCGG GACAGAGTGA CCCCCGTGGA GCTTCTCCGA GCCCGTGCTG AGGACCCTCT 5600
TGCAAAGGGC TCCACAGACC CCCGCCCTGG AGAGAGGAGT CTGAGCCTGG CTTAATAACA AACTGGGATG 5670
TGGCTGGGGG CGGACAGCGA CGGCGGGATT CAAAGACTTA ATTCCATGAG TAAATTCAAC CTTTCCACAT 5740
CCGAATGGAT TTGGATTTTA TCTTAATATT TTCTTAAATT TCATCAAATA ACATTCAGGA CTGCAGAAAT 5810
CCAAAGGCGT AAAACAGGAA CTGAGCTATG TTTGCCAAGG TCCAAGGACT TAATAACCAT GTTCAGAGGG 5880
ATTTTTCGCC CTAAGTACTT TTTATTGGTT TTCATAAGGT GGCTTAGGGT GCAAGGGAAA GTACACGAGG 5950
AGAGGCCTGG GCGGCAGGGC TATGAGCACG GCAGGGCCAC CGGGGAGAGA GTCCCCGGCC TGGGAGGCTG 6020
ACAGCAGGAC CACTGACCGT CCTCCCTGGG AGCTGCCACA TTGGGCAACG CGAAGGCGGC CACGCTGCGT 6090
GTGACTCAGG ACCCCATACC GGCTTCCTGG GCCCACCCAC ACTAACCCAG GAAGTCACGG AGCTCTGAAC 6160
CCGTGGAAAC GAACATGACC CTTGCCTGCC TGCTTCCCTG GGTGGGTCAA GGGTAATGAA GTGGTGTGCA 6230
GGAAATGGCC ATGTAAATTA CACGACTCTG CTGATGGGGA CCGTTCCTTC CATCATTATT CATCTTCACC 6300
CCCAAGGACT GAATGATTCC AGCAACTTCT TCGGGTGTGA CAAGCCATGA CAAAACTCAG TACAAACACC 6370
ACTCTTTTAC TAGGCCCACA GAGCACGGSC CACACCCCTG ATATATTAAG AGTCCAGGAG AGATGAGGCT 6440
GCTTTCAGCC ACCAGGCTGG GGTGACAACA GCGGCTGAAC AGTCTGTTCC TCTAGACTAG TAGACCCTGG 6510
CAGGCACTCC CCCAGATTCT AGGGCCTGGT TGCTGCTTCC CGAGGGCGCC ATCTGCCCTG GAGACTCAGC 6580
CTGGGGTGCC ACACTGAGGC CAGCCCTGTC TCCACACCCT CCGCCTCCAG GCCTCAGCTT CTCCAGCAGC 6650
TTCCTAAACC CTGGGTGGGC CGTGTTCCAG CGCTACTGTC TCACCTGTCC CACTGTGTCT TGTCTCAGCG 6720
ACGTAGCTCG CACGGTTCCT CCTCACATGG GGTGTCTGTC TCCTTCCCCA ACACTCACAT GCGTTGAAGG 6790
GAGGAGATTC TGCGCCTCCC AGACTGGCTC CTCTGAGCCT GAACCTGGCT CGTGGCCCCC GATGCAGGTT 6860
CCTGGCGTCC GGCTGCACGC TGACCTCCAT TTCCAGGCGC TCCCCGTCTC CTGTCATCTG CCGGGGCCTG 6930
CCGGTGTGTT CTTCTGTTTC TGTGCTCCTT TCCACGTCCA GCTGCGTGTG TCTCTGCCCG CTAGGGTCTC 7000
GGGGTTTTTA TAGGCATAGG ACGGGGGCGT GGTGGGCCAG GGCGCTCTTG GGAAATGCAA CATTTGGGTG 7070
TGAAAGTAGG AGTGCCTGTC CTCACCTAGG TCCACGGGCA CAGGCCTGGG GATGGAGCCC CCGCCAGGGA 7140
CCCGCCCTTC TCTGCCCAGC ACTTTCCTGC CCCCCTCCCT CTGGAACACA GAGTGGCAGT TTCCACAAGC 7210
ACTAAGCATC CTCTTCCCAA AAGACCCAGC ATTGGCACCC CTGGACATTT GCCCCACAGC CCTGGGAATT 7280
CACGTGACTA CGCACATCAT GTACACACTC CCGTCCACGA CCGACCCCCG CTGTTTTATT TTAATAGCTA 7350
CAAAGCAGGG AAATCCCTGC TAAAATGTCC TTTAACAAAC TGGTTAAACA AACGGGTCCA TCCGCACGGT 7420
GGACAGTTCC TCACAGTGAA GAGGAACATG CCGTTTATAA AGCCTGCAGG CATCTCAAGG GAATTACGCT 7490
GAGTCAAAAC TGCCACCTCC ATGGGATACG TACGCAACAT GCTCAAAAAG AAAGAATTTC ACCCCATGGC 7560
AGGGGAGTGG TTAGGGGGGT TAAGGACGGT GGGGGCGGCA GCTGGGGGCT ACTGCACGCA CCTTTTACTA 7630
AAGCCAGTTT CCTGGTTCTG ATGGTATTGG CTCAGTTATG GGAGACTAAC CATAGGGGAG TGGGGATGGG 7700
GGAACCCGGA GGCTGTGCCA TCTTTGCCAT GCCCGAGTGT CCTGGGCAGG ATAATGCTCT AGAGATGCCC 7770
ACGTCCTGAT TCCCCCAAAC CTGTGGACAG AACCCGCCCG GCCCCAGGGC CTTTGCAGGT GTGATCTCCG 7840
TGAGGACCCT GAGGTCTGGG ATCCTTCGGG ACTACCTGCA GGCCCGAAAA GTAATCCAGG GGTTCTGGGA 7910
AGAGGCGGGC AGGAGGGTCA GAGGGGGGCA GCCTCAGGAC GATGGAGGCA GTCAGTCTGA GGCTGAAAAG 7980
GGAGGGAGGG CCTCGAGCCC AGGCCTGCAA GCGCCTCCAG AAGCTGGAAA AAGCGGGGAA GGGACCCTCC 8050
ACGGAGCCTG CAGCAGGAAG GCACGGCTGG CCCTTAGCCC ACCAGGGCCC ATCGTGGACC TCCGGCCTCC 8120
GTGCCATAGG AGGGCACTCG CGCTGCCCTT CTAGCATGAA GTGTGTGGGG ATTTGCAGAA GCAACAGGAA 8190
ACCCATGCAC TGTGAATCTA GGATTATTTC AAAACAAAGG TTTACAGAAA CATCCAAGGA CAGGGCTGAA 8260
GTGCCTCCGG GCAAGGGCAG GGCAGGCACG AGTGATTTTA TTTAGCTATT TTATTTTATT TACTTACTTT 8330
CTGAGACAGA GTTATGCTCT TGTTGCCCAG GCTGGAGTGC AGCGGCATGA TCTTGGCTCA CTGCAACCTC 8400
CGTCTCCTGG GTTCAAGCAA TTCTCGTGCC TCAGCCTCCC AAGTAGCTGG GATTTCAGGC GTGCACCACC 8470
ACACCCGGCT AATTTTGTAT TTTTAGTAGA GATGGGCTTT CACCATGTTG GTCAAGCTGA TCTCAAAATC 8540
CTGACCTCAG GTGATCCGCC CACCTCAGCC TCCCAAAGTG CTGGGATTAC AGGCATGAGC CACTGCACCT 8610
GGCCTATTTA ACCATTTTAA AACTTCCCTG GGCTCAAGTC ACACCCACTG GTAAGGAGTT CATGGAGTTC 8680
AATTTCCCCT TTACTCAGGA GTTACCCTCC TTTGATATTT TCTGTAATTC TTCGTAGACT GGGGATACAC 8750
CGTCTCTTGA CATATTCACA GTTTCTGTGA CCACCTGTTA TCCCATGGGA CCCACTGCAG GGGCAGCTGG 8820
GAGGCTGCAG GCTTCAGGTC CCAGTGGGGT TGCCATCTGC CAGTAGAAAC CTGATGTAGA ATCAGGGCGC 8890
AAGTGTGGAC ACTGTCCTGA ATCTCAATGT CTCAGTGTGT GCTGAAACAT GTAGAAATTA AAGTCCATCC 8960
CTCCTACTCT ACTGGGATTG AGCCCCTTCC CTATCCCCCC CCAGGGGCAG AGGAGTTCCT CTCACTCCTG 9030
TGGAGGAAGG AATGATACTT TGTTATTTTT CACTGCTGGT ACTGAATCCA CTGTTTCATT TGTTGGTTTG 9100
TTTGTTTTGT TTTGAGAGGC GGTTTCACTC TTGTTGCTCA GGCTGGAGGG AGTGCAATGG CGCGATCTTG 9170
GCTTACTGCA GCCTCTGCCT CCCAGGTTCA AGTGATTCTC CTGCTTCCGC CTCCCATTTG GCTGGGATTA 9240
CAGGCACCCG CCACCATGCC CAGCTAATTT TTTGTATTTT TAGTAGAGAC GGGGGTGGGT GGGGTTCACC 9310
ATGTTGGCCA GGCTGGTCTC GAACTTCTGA CCTCAGATGA TCCACCTGCC TCTGCCTCCT AAAGTGCTGG 9380
GATTACAGGT GTGAGCCACC ATGCCCAGCT CAGAATTTAC TCTGTTTAGA AACATCTGGG TCTGAGGTAG 9450
GAAGCTCACC CCACTCAAGT GTTGTGGTGT TTTAAGCCAA TGATAGAATT TTTTTATTGT TGTTAGAACA 9520
CTCTTGATGT TTTACACTGT GATGACTAAG ACATCATCAG CTTTTCAAAG ACACACTAAC TGCACCCATA 9590
ATACTGGGGT GTCTTCTGGG TATCAGCAAT CTTCATTGAA TGCCGGGAGG CGTTTCCTCG CCATGCACAT 9660
GGTGTTAATT ACTCCAGCAT AATCTTCTGC TTCCATTTCT TCTCTTCCCT CTTTTAAAAT TGTGTTTTCT 9730
ATGTTGGCTT CTCTGCAGAG AACCAGTGTA AGCTACAACT TAACTTTTGT TGGAACAAAT TTTCCAAACC 9800
GCCCCTTTGC CCTAGTGGCA GAGACAATTC ACAAACACAG CCCTTTAAAA AGGCTTAGGG ATCACTAAGG 9870
GGATTTCTAG AAGAGCGACC TGTAATCCTA AGTATTTACA AGACGAGGCT AACCTCCAGC GAGCGTGACA 9940
GCCCAGGGAG GGTGCGAGGC CTGTTCAAAT GCTAGCTCCA TAAATAAAGC AATTTCCTCC GGCAGTTTCT 10010
GAAAGTAGGA AAGGTTACAT TTAAGGTTGC GTTTGTTAGC ATTTCAGTGT TTGCCGACCT CAGCTACAGC 10080
ATCCCTGCAA GGCCTCGGGA GACCCAGAAG TTTCTCGCCC CCTTAGATCC AAACTTGAGC AACCCGGAGT 10150
CTGGATTCCT GGGAAGTCCT CAGCTGTCCT GCGGTTGTGC CGGGGCCCCA GGTCTGGAGG GGACCAGTGG 10220
CCGTGTGGCT TCTACTGCTG GGCTGGAAGT CGGGCCTCCT AGCTCTGCAG TCCGAGGCTT GGAGCCAGGT 10290
GCCTGGACCC CGAGGCTGCC CTCCACCCTG TGCGGGCGGG ATGTGACCAG ATGTTGGCCT CATCTGCCAG 10360
ACAGAGTGCC GGGGCCCAGG GTCAAGGCCG TTGTGGCTGG TGTGAGGCGC CCGGTGCGCG GCCAGCAGGA 10430
GCGCCTGGCT CCATTTCCCA CCCTTTCTCG ACGGGACCGC CCCGGTGGGT GATTAACAGA TTTGGGGTGG 10500

TTTGCTCATG GTGGGGACCC CTCGCCGCCT GAGAACCTGC AAAGAGAAAT GACGGGCCTG TGTCAAGGAG 10570
CCCAAGTCGC GGGGAAGTGT TGCAGGGAGG CACTCCGGGA GGTCCCGCGT GCCCGTCCAG GGAGCAATGC 10640
GTCCTCGGGT TCGTCCCCAG CCGCGTCTAC GCGCCTCCGT CCTCCCCTTC ACGTCCGGCA TTCGTGGTGC 10710
CCGGAGCCCG ACGCCCCGCG TCCGGACCTG GAGGCAGCCC TGGGTCTCCG GATCAGGCCA GCGGCCAAAG 10780
GGTCGCCGCA CGCACCTGTT CCCAGGGCCT CCACATCATG GCCCCTCCCT CGGGTTACCC CACAGCCTAG 10850
GCCGATTCGA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCCT GGGAGCGCGA GCGGCGCGCG 10920
GGCGGGGAAG CGCGGCCCAG ACCCCCGGGT CCGCCCGGAG CAGCTGCGCT GTCGGGGCCA GGCCGGGCTC 10990
CCAGTGGATT CGCGGGCACA GACGCCCAGG ACCGCGCTCC CCACGTGGCG GAGGGACTGG GGACCCGGGC 11060
ACCCGTCCTG CCCCTTCACC TTCCAGCTCC GCCTCCTCCG CGCGGACCCC GCCCCGTCCC GACCCCTCCC 11130
GGGTCCCCGG CCCAGCCCCC TCCGGGCCCT CCCAGCCCCT CCCCTTCCTT TCCGCGGCCC CGCCCTCTCC 11200
TCGCGGCGCG AGTTTCAGGC AGCGCTGCGT CCTGCTGCGC ACGTGGGAAG CCCTGGCCCC GGCCACCCCC 11270
GCGATGCCGC GCGCTCCCCG CTGCCGAGCC GTGCGCTCCC TGCTGCGCAG CCACTACCGC GAGGTGCTGC 11340
CGCTGGCCAC GTTCGTGCGG CGCCTGGGGC CCCAGGGCTG GCGGCTGGTG CAGCGCGGGG ACCCGGCGGC 11410
TTTCCGCGCG CTGGTGGCCC AGTGCCTGGT GTGCGTGCCC TGGGACGCAC GGCCGCCCCC CGCCGCCCCC 11480
TCCTTCCGCC AGGTGGGCCT CCCCGGGGTC GGCGTCCGGC TGGGGTTGAG GGCGGCCGGG GGGAACCAGC 11550
GACATGCGGA GAGCAGCGCA GGCGACTCAG GGCGCTTCCC CCGCAGGTGT CCTGCCTGAA GGAGCTGGTG 11620
GCCCGAGTGC TGCAGAGGCT GTGCGAGCGC GGCGCGAAGA ACGTGCTGGC CTTCGGCTTC GCGCTGCTGG 11690
ACGGGGCCCG CGGGGGCCCC CCCGAGGCCT TCACCACCAG CGTGCGCAGC TACCTGCCCA ACACGGTGAC 11760
CGACGCACTG CGGGGGAGCG GGGCGTGGGG GCTGCTGCTG CGCCGCGTGG GCGACGACGT GCTGGTTCAC 11830
CTGCTGGCAC GCTGCGCGCT CTTTGTGCTG GTGGCTCCCA GCTGCGCCTA CCAGGTGTGC GGGCCGCCGC 11900
TGTACCAGCT CGGCGCTGCC ACTCAGGCCC GGCCCCCGCC ACACGCTAGT GGACCCCGAA GGCGTCTGGG 11970
ATGCGAACGG GCCTGGAACC ATAGCGTCAG GGAGGCCGGG GTCCCCCTGG GCCTGCCAGC CCCGGGTGCG 12040
AGGAGGCGCG GGGGCAGTGC CAGCCGAAGT CTGCCGTTGC CCAAGAGGCC CAGGCGTGGC GCTGCCCCTG 12110
AGCCGGAGCG GACGCCCGTT GGGCAGGGGT CCTGGGCCCA CCCGGGCAGG ACGCGTGGAC CGAGTGACCG 12180
TGGTTTCTGT GTGGTGTCAC CTGCCAGACC CGCCGAAGAA GCCACCTCTT TGGAGGGTGC GCTCTCTGGC 12250
ACGCGCCACT CCCACCCATC CGTGGGCCGC CAGCACCACG CAGGCCCCCC ATCCACATCG CGGCCACCAC 12320
GTCCCTGGGA CACGCCTTGT CCCCCGGTGT ACGCCGAGAC CAAGCACTTC CTCTACTCCT CAGGCGACAA 12390
GGAGCAGCTG CGGCCCTCCT TCCTACTCAG CTCTCTGAGG CCCAGCCTGA CTGGCGCTCG GAGGCTCGTG 12460
GAGACCATCT TTCTGGGTTC CAGGCCCTGG ATGCCAGGGA CTCCCCGCAG GTTGCCCCGC CTGCCCCAGC 12530
GCTACTGGCA AATGCGGCCC CTGTTTCTGG AGCTGCTTGG GAACCACGCG CAGTGCCCCT ACGGGGTGCT 12600
CCTCAAGACG CACTGCCCGC TGCGAGCTGC GGTCACCCCA GCAGCCGGTG TCTGTGCCCG GGAGAAGCCC 12670
CAGGGCTCTG TGGCGGCCCC CGAGGAGGAG GACACAGACC CCCGTCGCCT GGTGCAGCTG CTCCGCCAGC 12740
ACAGCAGCCC CTGGCAGGTG TACGGCTTCG TGCGGGCCTG CCTGCGCCGG CTGGTGCCCC CAGGCCTCTG 12810
GGGCTCCAGG CACAACGAAC GCCGCTTCCT CAGGAACACC AAGAAGTTCA TCTCCCTGGG GAAGCATGCC 12880
AAGCTCTCGC TGCAGGAGCT GACGTGGAAG ATGAGCGTGC GGGACTGCGC TTGGCTGCGC AGGAGCCCAG 12950
GTGAGGAGGT GGTGGCCGTC GAGGGCCCAG GCCCCAGAGC TGAATGCAGT AGGGGCTCAG AAAAGGGGGC 13020
AGGCAGAGCC CTGGTCCTCC TGTCTCCATC GTCACGTGGG CACACGTGGC TTTTCGCTCA GGACGTCGAG 13090
TGGACACGGT GATCTCTGCC TCTGCTCTCC CTCCTGTCCA GTTTGCATAA ACTTACGAGG TTCACCTTCA 13160
CGTTTTGATG GACACGCGGT TTCCAGGCGC CGAGGCCAGA GCAGTGAACA GAGGAGGCTG GGCGCGGCAG 13230
TGGAGCCGGG TTGCCGGCAA TGGGGAGAAG TGTCTGGAAG CACAGACGCT CTGGCGAGGG TGCCTGCAGG 13300
TTACCTATAA TCCTCTTCGC AATTTCAAGG GTGGGAATGA GAGGTGGGGA CGAGAACCCC CTCTTCCTGG 13370
GGGTGGGAGG TAAGGGTTTT GCAGGTGCAC GTGGTCAGCC AATATGCAGG TTTGTGTTTA AGATTTAATT 13440
GTGTGTTGAC GGCCAGGTGC GGTGGCTCAC GCCGGTAATC CCAGCACTTT GGGAAGCTGA GGCAGGTGGA 13510
TCACCTGAGG TCAGGAGTTT GAGACCAGCC TGACCAACAT GGTGAAACCC TATCTGTACT AAAAATACAA 13580
AAATTAGCTG GGCATGGTGG TGTGTGCCTG TAATCCCAGC TACTTGGGAG GCTGAGGCAG GAGAATCACT 13650
TGAACCCAGG AGGCGGAGGC TGCAGTGAGC TGAGATTGTG CCATTGTACT CCAGCCTGGG CGACAAGAGT 13720
GAAACTCTGT CTTTAAAAAA AAAAAGTGTT CGTTGATTGT GCCAGGACAG GGTAGAGGGA GGGAGATAAG 13790
ACTGTTCTCC AGCACAGATC CTGGTCCCAT CTTTAGGTAT GAAGAGGGCC ACATGGGAGC AGAGGACAGC 13860
AGATGGCTCC ACCTGCTGAG GAAGGGACAG TGTTTGTGGG TGTTCAGGGG ATGGTGCTGC TGGGCCCTGC 13930
CGTGTCCCCA CCCTGTTTTT CTGGATTTGA TGTTGAGGAA CCTCCGCTCC AGCCCCCTTT TGGCTCCCAG 14000
TGCTCCCAGG CCCTACCGTG GCAGCTAGAA GAAGTCCCGA TTTCACCCCC TCCCCACAAA CTCCCAAGAC 14070
ATGTAAGACT TCCGGCCATG CAGACAAGGA GGGTGACCTT CTTGGGGCTC TTTTTTTTCT TTTTTCTTT 14140
TTATGGTGGC AAAAGTCATA TAACATGAGA TTGGCACTCC TAACACCGTT TTCTGTGTAC AGTGCAGAAT 14210
TGCTAACTCG GCGGTGTTTA CAGCAGGTTG CTTGAAATGC TGCGTCTTGC GTGACTGGAA GTCCCTACCC 14280
ATCGAACGGC AGCTGCCTCA CACCTGCTGC GGCTCAGGTG GACCACGCCG AGTCAGATAA GCGTCATGCA 14350
ACCCAGTTTT GCTTTTTGTG CTCCAGCTTC CTTCGTTGAG GAGAGTTTGA GTTCTCTGAT CAGGACTCTG 14420
CCTGTCATTG CTGTTCTCTG ACTTCAGATG AGGTCACAAT CTGCCCCTGG CTTATGCAGG GAGTGAGGCG 14490
TGGTCCCCGG GTGTCCCTGT CACGTGCAGG GTGAGTGAGG CGTTGCCCCC AGGTGTCCCT GTCACGTGTA 14560
GGGTGAGTGA GGCGCGGCCC CCGGGTGTCC CTGTCCCGTG CAGCGTGATT GAGGTGTGGC CCCCGGGTGT 14630
CCCTGTCACG TGTAGGGTGA GTGAGGCGCC ATCCCCGGGT GTCCCTGTCA CGTGTAGGGT GAGTGAGGCG 14700
TGGTCCCCGG GTGTCCCTGT CCCGTGCAGG GTGAGTGAGG CACTGTCCCC GGGTGTCCCT GTCACGTGCA 14770
GGGTGAGTGA GGCGCGGTCC CCGGGTGTCC CTCTCAGGTG TAGGGTGAGT GAGGCGCGGC CCCAGGGTGT 14840
CCCTGTCACG TGTAGGGTGA GTGAGGCACC GTCCCTGGGT GTCCCTCCCA GGTATAGGGT GAGTGAGGCA 14910
CTGTCCCCGG GTGTCCCTGT CACGTGCAGG GTGAGTGAGG CGCGGCCCCC GGGTGTCCCT CTCAGGTGCA 14980
GGGTGAGTGA GGCGCTGTCC CTGGGTGTCC CTGTCTCGTG TAGGGTGAGT GAGGCTCTGT CCCCAGGTGT 15050
CCTTGGCGTT TGCTCACTTG AGCTTGCTCC TGAATGTTTG CTCTTTCTAT AGCCACAGCT GCGCCGGTTG 15120
CCCATTGCCT GGGTAGATGG TGCAGGCGCA GTGCTGGTCC CCAAGCCTAT CTTTTCTGAT GCTCGGCTCT 15190
TCTTGGTCAC CTCTCCGTTC CATTTTGCTA CGGGGACACG GGACTGCAGG CTCTCGCCTC CCGCGTGCCA 15260
GGCACTGCAG CCACAGCTTC AGGTCCGCTT GCCTCTGTTG GGCCTGGCTT GCTCACCACG TGCCCGCCAC 15330
ATGCATGCTG CCAATACTCC TCTCCCAGCT TGTCTCATGC CGAGGCTGGA CTCTGGGCTG CCTGTGTCTG 15400
CTGCCACGTG TTGCTGGAGA CATCCCAGAA AGGGTTCTCT GTGCCCTGAA GGAAAGCAAG TCACCCCAGC 15470
CCCCTCACTT GTCCTGTTTT CTCCCAAGCT GCCCCTCTGC TTGGCCCCCT TGGGTGGGTG GCAACGCTTG 15540
TCACCTTATT CTGGGCACCT GCCGCTCATT GCTTAGGCTG GGCTCTGCCT CCAGTCGCCC CCTCACATGG 15610
ATTGACGTCC AGCCACAGGT TGGAGTGTCT CTGTCTGTCT CCTGCTCTGA GACCCACGTG GAGGGCCGGT 15680
GTCTCCGCCA GCCTTCGTCA GACTTCCCTC TTGGGTCTTA GTTTTGAATT TCACTGATTT ACCTCTGACG 15750
TTTCTATCTC TCCATTGTAT GCTTTTTCTT GGTTTATTCT TTCATTCCTT TTCTAGCTTC TTAGTTTAGT 15820
CATGCCTTTC CCTCTAAGTG CTGCCTTACC TGCACCCTGT GTTTTGATGT GAAGTAATCT CAACATCAGC 15890

CACTTTCAAG TGTTCTTAAA ATACTTCAAA GTGTTAATAC TTCTTTTAAG TATTCTTATT CTGTGATTTT 15960
TTTCTTTGTG CACGCTGTGT TTTGACGTGA AATCATTTTG ATATCAGTGA CTTTTAAGTA TTCTTTAGCT 16030
TATTCTGTGA TTTCTTTGAG CAGTGAGTTA TTTGAACACT GTTTATGTTC AAGATATGTA GAGTATCAAG 16100
ATACGTAGAG TATTTTAAGT TATCATTTTA TTATTGATTT CTAACTCAGT TGTGTAGTGG TCTGTATAAT 16170
ACCAATTATT TGAAGTTTGC GGAGCCTTGC TTTGTGATCT AGTGTGTGCA TGGTTTCCAG AACTGTCCAT 16240
TGTAAATTTG ACATCCTGTC AATAGTGGGC ATGCATGTTC ACTATATCCA GCTTATTAAG GTCCAGTGCA 16310
AAGCTTCTGT CTCCTTCTAG ATGCATGAAA TTCCAAGAAG GAGGCCATAG TCCCTCACCT GGGGGATGGG 16380
TCTGTTCATT TCTTCTCGTT TGGTAGCATT TATGTGAGGC ATTGTTAGGT GCATGCACGT GGTAGAATTT 16450
TTATCTTCCT GATGAGTGAA TCTTTTGGAG ACTTCTATGT CTCTAGTAAT CTAGTAATTC TTTTTTTAAA 16520
TTGCTCTTAG TACTGCCACA CTGGGCTTCT TTTGATTAGT ATTTTCCTGC TGTGTCTGTT TTCTGCCTTT 16590
AATTTATATA TATATATATA TTTTTTTTTT TTTTGAGACA GAGTCTTGGT CTGTCGCCCA GGGTGAGTGC 16660
AGTGGTGTGA TCACAGGTCA GTGTAACTTT TACCTTCTGG CCTGAGCCGT CCTCTCACCT CAGCCTCCTG 16730
AGTAGCTGGA ACTGCAGACA CGCACCGCTA CACCTGGCTA ATTTTTAAAT TTTTTCTGGA GACAGGGTCT 16800
TGCTGTGTTG CCCAGGCTGG TCTCAAACTC TTGGACTCAA GGGATCCATC TACCTCGGCT TCCCAAAGTG 16870
CTGAATTACA GGCATGAGCC ACCATGTCTG GCCTAATTTT CAACACTTTT ATATTCTTAT AGTGTGGGTA 16940
TGTCCTGTTA ACAGCATGTA GGTGAATTTC CAATCCAGTC TGACAGTCGT TGTTTAACTG GATAACCTGA 17010
TTTATTTTCA TTTTTTTGTC ACTAGAGACC CGCCTGGTGC ACTCTGATTC TCCACTTGCC TGTTGCATGT 17080
CCTCGTTCCC TTGTTTCTCA CCACCTCTTG GGTTGCCATG TGCGTTTCCT GCCGAGTGTG TGTTGATCCT 17150
CTCGTTGCCT CCTGGTCACT GGGCATTTGC TTTTATTTCT CTTTGCTTAG TGTTACCCCC TGATCTTTTT 17220
ATTGTCGTTG TTTGCTTTTG TTTATTGAGA CAGTCTCACT CTGTCACCCA GGCTGGAGTG TAATGGCACA 17290
ATCTCGGCTC ACTGCAACCT CTGCCTCCTC GGTTCAAGCA GTTCTCATTC CTCAACCTCA TGAGTAGCTG 17360
GGATTACAGG CGCCCACCAC CACGCCTGGC TAATTTTTGT ATTTTTAGTA GAGATAGGCT TTCACCATGT 17430
TGGCCAGGCT GGTCTCAAAC TCCTGACCTC AAGTGATCTG CCCGCCTTGG CCTCCCACAG TGCTGGGATT 17500
ACAGGTGCAA GCCACCGTGC CCGGCATACC TTGATCTTTT AAAATGAAGT CTGAAACATT GCTACCCTTG 17570
TCCTGAGCAA TAAGACCCTT AGTGTATTTT AGCTCTGGCC ACCCCCCAGC CTGTGTGCTG TTTTCCCTGC 17640
TGACTTAGTT CTATCTCAGG CATCTTGACA CCCCCACAAG CTAAGCATTA TTAATATTGT TTTCCGTGTT 17710
GAGTGTTTCT GTAGCTTTGC CCCCGCCCTG CTTTTCCTCC TTTGTTCCCC GTCTGTCTTC TGTCTCAGGC 17780
CCGCCGTCTG GGGTCCCCTT CCTTGTCCTT TGCGTGGTTC TTCTGTCTTG TTATTGCTGG TAAACCCCAG 17850
CTTTACCTGT GCTGGCCTCC ATGGCATCTA GCGACGTCCG GGGACCTCTG CTTATGATGC ACAGATGAAG 17920
ATGTGGAGAC TCACGAGGAG GGCGGTCATC TTGGCCCGTG AGTGTCTGGA GCACCACGTG GCCAGCGTTC 17990
CTTAGCCAGT GAGTGACAGC AACGTCCGCT CGGCCTGGGT TCAGCCTGGA AAACCCCAGG CATGTCGGGG 18060
TCTGGTGGCT CCGCGGTGTC GAGTTTGAAA TCGCGCAAAC CTGCGGTGTG GCGCCAGCTC TGACGGTGCT 18130
GCCTGGCGGG GGAGTGTCTG CTTCCTCCCT TCTGCTTGGG AACCAGGACA AAGGATGAGG CTCCGAGCCG 18200
TTGTCGCCCA ACAGGAGCAT GACGTGAGCC ATGTGGATAA TTTTAAAATT TCTAGGCTGG GCGCGGTGGC 18270
TCACGCCTGT AATCCCAGCA CTTTGGGAGG CCAAGGCGGG TGGATCACGA GGTCAGGAGG TCGAGACCAT 18340
CCTGGCCAAC ATGATGAAAC CCCATCTGTA CTAAAAACAC AAAAATTAGC TGGGCGTGGT GGCGGGTGCC 18410
TGTAATCCCA GCTACTCGGG AGGCTGAGGC AGGAGAATTG CTTGAACCTG GGAGTTGGAA GTTGCAGTGA 18480
GCCGACATTG CACCACTGCA CTCCAGCCTG GCAACACAGC GAGACTCTGT CTCAAAAAAA AAAAAAAAAA 18550
AAAAAAAAAA AATTCTAGTA GCCACATTAA AAAAGTAAAA AAGAAAAGGT GAAATTAATG TAATAATAGA 18620
TTTTACTGAA GCCCAGCATG TCCACACCTC ATCATTTTAG GGTGTTATTG GTGGGAGCAT CACTCACAGG 18690
ACATTTGACA TTTTTTGAGC TTTGTCTGCG GGATCCCGTG TGTAGGTCCC GTGCGTGGCC ATCTCGGCCT 18760
GGACCTGCTG GGCTTCCCAT GGCCATGGCT GTTGTACCAG ATGGTGCAGG TCCGGGATGA GGTCGCCAGG 18830
CCCTCAGTGA GCTGGATGTG CAGTGTCCGG ATGGTGCACG TCTGGGATGA GGTCGCCAGG CCCTGCTGTG 18900
AGCTGGATGT GTGGTGTCTG GATGGTGCAG GTCAGGGGTG AGGTCTCCAG GCCCTCGGTG AGCTGGAGGT 18970
ATGGAGTCCG GATGATGCAG GTCCGGGGTG AGGTCGCCAG GCCCTGCTGT GAGCTGGATG TGTGGTGTCT 19040
GGATGGTGCA GGTCAGGGGT GAGGTCTCCA GGCCCTCGGT AAGCTGGAGG TATGGAGTCC GGATGATGCA 19110
GGTCCGGGGT GAGGTCGCCA GGCCCTGCTG TGAGCTGGAT GTGTGGTGTC TGGATGGTGC AGGTCTGGGG 19180
TGAGGTCACC AGGCCCTGCG GTGAGCTGGG TGTGCGGTGT CTGGATGGTG CAGGTCTGGA GTGAGGTCGC 19250
CAGACGGTGC CAGACCATGC GGTGAGCTGG ATATGCGGTG TCCGGATGGT GCAGGTCTGG GGTGAGGTTG 19320
CCAGGCCCTG CTGTGAGTTG GATGTGGGGT GTCCGGATGC TGCAGGTCCG GTGTGAGGTC ACCAGGCCCT 19390
GCTGTGAGCT GGATGTGTGG TGTCTGGATG GTGCAGGTCT GGGGTGAAGG TCGCCAGGCC CCTGCTTGTG 19460
AGCTGGATGT GTGGTGTCTG GATGGTGCAG GTCTGGAGTG AGGTCGCCAG GCCCTCGGTG AGCTGGATGT 19530
GCAGTGTCCA GATGGTGCAG GTCCGGGGTG AGGTCGCCAG ACCCTGCGGT GAGCTGGATG TGCGGTGTCT 19600
GGATGGTGCA GGTCTGGAGT GAGGTCGCCA GGCCCTCGGT GAGCTGGATG TATGGAGTCC GGATGGTGCC 19670
GGTCCGGGGT GAGGTCGCCA GACCCTGCTG TGAGCTGGAT GTGCGGTGTC TGGATGGTAC AGGTCTGGAG 19740
TGAGGTCGCC AGACCCTGCT GTGAGCTGGA TATGCGGTGT CCGGATGGTG CAGGTCAGGG GTGAGGTCTC 19810
CAGGCCCTCG GTGAGCTGGA GGTATGGAGT CCGGATGATG CAGGTCCGGG GTGAGGTCGC CAGGCCCTGC 19880
TGTGAACTGG ATGTGCGGCG TCTGGATGGT GCAGGTCTGG GGTGTGGTCG CCAGGCCCTC GGTGAGCTGG 19950
AGGTATGGAG TCCGGATGAT GCAGGTCCGG GGTGAGGTCG CCAGGCCCTG CTGTGAGCTG GATGTGCGGC 20020
GTCTGGATGG TGCAGGTCTG GGGTGTGGTC GCCAGGCCCT CGGTGAGCTG GAGGTATGGA GTCCGGATGA 20090
TGCAGGTCCG GGGTGAGGTT GCCAGGCCCT GCTGTGAGCT GGATGTGCTG TATCCGGATG GTGCAGTCCG 20160
GGGTGAGGTC GCCAGGCCCT GCTGTGAGCT GGATGTGCTG TATCCGGATG GTGCAGGTCT GGGGTGAGGT 20230
CACCAGGCCC TGCGGTGAGC TGGTTGTGCG GTGTCCGGTT GCTGCAGGTC CGGGGTGAGT TCGCCAGGCC 20300
CTCGGTGAGC TGGATGTGCG GTGTCCCCGT GTCCGGATGG TGCAGGTCCA GGGTGAGGTC GCTAGGCCCT 20370
TGGTGGGCTG GATGTGCCGT GTCCGGATGG TGCAGGTCTG GGGTGAGGTC GCCAGGCCCT TGGTGAGCTG 20440
GATGTGCGGT GTCTGCATGG TGCAGGTCTG GGGTGAGGTC GCCAGGCCCT TGGTGGGCTG GATGTGTGGT 20510
GTCCGGATGG TGCAGGTCCG GCGTGAGGTC GCCAGGCCCT GCTGTGAGCT GGATGTGCGG TGTCTGGATG 20580
GTGCAGGTCC GGGGTGAGGT AGCCAAGGCC TTCGGTGAGC TGGATGTGGG GTGTCCGGAT GGTGCAGGTC 20650
CGGGGTGAGG TCGCCAGGCC CTGCGGTTAG CTGGATATGC GGTGTCCGGA TGGTGCAGGT CCGGGGTGAG 20720
GTCACCAGGC CCTGCGGTTA GCTGGATGTG CGGTGTCTGG ATGGTGCAGG TCCGGGGTGA GGTCGCCAGG 20790
CCCTGCTGTG AGCTGGATGT GCTGTATCCG GATGGTGCAG GTCCGGGGTG AGGTCGCCAG GCCCTGCAGT 20860
GAGCTGGATG TGCTGTATCC GGATGGTGCA GGTCTGGCGT GAGGTCGCCA GGCCCTGCGG TTAGCTGGAT 20930
ATGCGGTGTC GGATGGTGCA GGTCCGGGGT GAGGTCACCA GGCCCTGCGG TTAGCTGGAT GTGCGGTGTC 21000
CGGATGGTGC AGGTCTGGGG TGAGGTCGCC AGGCCCTGCT GTGAGCTGGA TGTGCTGTAT CCGGATGGTG 21070
CAGGTCCGGG GTGAGGTCGC CAGGCCCTGC GGTGAGCTGG ATGTGCTGTA TCCGGATGGT GCAGGTCTGG 21140
CGTGAGGTCG CCAGGCCCTG CGGTGAGCTG GATGTGCAGT GTACGGATGG TGCAGGTCCG GGGTGAGGTC 21210
GCCAGGCCCT GCGGTGGGCT GTATGTGTGT TGTCTGGATG GTGCAGGTCC GGGGTGAGTT CGCCAGGCCC 21280

TGCGGTGAGC TGGATGTGTG GTGTCTGGAT GCTGCAGGTC CGGGGTGAGT TCGCCAGGCC CTCGGTGAGC 21350
TGGATATGCG GTGTCCCCGT GTCCGAATGG TGCAGGTCCA GGGTGAGGTC GCCAGGCCCT TGGTGGGCTG 21420
GATGTGCCGT GTCCGGATGG TGCAGGTCTG GGGTGAGGTC GCCAGGCCCT TGGTGAGCTG GATGTGCGGT 21490
GTCCGGATGG TGCAGGTCCG GGGTGAGGTC ACCAGGCCCT CGGTGATCTG GATGTGGCAT GTCCTTCTCG 21560
TTTAAGGGGT TGGCTGTGTT CCGGCCGCAG AGCACCGTCT GCGTGAGGAG ATCCTGGCCA AGTTCCTGCA 21630
CTGGCTGATG AGTGTGTACG TCGTCGAGCT GCTCAGGTCT TTCTTTTATG TCACGGAGAC CACGTTTCAA 21700
AAGAACAGGC TCTTTTTCTA CCGGAAGAGT GTCTGGAGCA AGTTGCAAAG CATTGGAATC AGGTACTGTA 21770
TCCCCACGCC AGGCCTCTGC TTCTCGAAGT CCTGGAACAC CAGCCCGGCC TCAGCATGCG CCTGTCTCCA 21840
CTTGCCTGTG CTTCCCTGGC TGTGCAGCTC TGGGCTGGGA GCCAGGGGCC CCGTCACAGG CCTGGTCCAA 21910
GTGGATTCTG TGCAAGGCTC TGACTGCCTG GAGCTCACGT TCTCTTACTT GTAAAATCAG GAGTTTGTGC 21980
CAAGTGGTCT CTAGGGTTTG TAAAGCAGAA GGGATTTAAA TTAGATGGAA ACACTACCAC TAGCCTCCTT 22050
GCCTTTCCCT GGGATGTGGG TCTGATTCTC TCTCTCTTTT TTTTTCTTT TTTGAGATGG AGTCTCACTC 22120
TGTTGCCCAG GCTGGAGTGC AGTGGCATAA TCTTGGCTCA CTGCAACCTC CACCTCCTGG GTTTAAGCGA 22190
TTCACCAGCC TCAGCCTCCT AAGTAGCTGG GATTACAGGC ACCTGCCACC ACGCCTGGCT AATTTTTGTA 22260
CTTTTAGGAG AGACGGGGTT TCACCATGTT GGCCAGGCTG GTCTCGAACT CATGACCTCA GGTGATCCAC 22330
CCACCTTGGC CTCCCAAAGT GCTGGGTTTA CAGGCTAAGC CACCGTGCCC AGCCCCCGAT TCTCTTTTAA 22400
TTCATGCTGT TCTGTATGAA TCTTCAATCT ATTGGATTTA GGTCATGAGA GGATAAAATC CCACCCACTT 22470
GGCGACTCAC TGCAGGGAGC ACCTGTGCAG GGAGCACCTG GGGATAGGAG AGTTCCACCA TGAGCTAACT 22540
TCTAGGTGGC TGCATTTGAA TGGCTGTGAG ATTTTGTCTG CAATGTTCGG CTGATGAGAG TGTGAGATTG 22610
TGACAGATTC AAGCTGGATT TGCATCAGTG AGGGACGGGA GCGCTGGTCT GGGAGATGCC AGCCTGGCTG 22680
AGCCCAGGCC ATGGTATTAG CTTCTCCGTG TCCCGCCCAG GCTGACTGTG GAGGGCTTTA GTCAGAAGAT 22750
CAGGGCTTCC CCAGCTCCCC TGCACACTCG AGTCCCTGGG GGGCCTTGTG ACACCCCATG CCCCAAATCA 22820
GGATGTCTGC AGAGGGAGCT GGCAGCAGAC CTCGTCAGAG GTAACACAGC CTCTGGGCTG GGGACCCCGA 22890
CGTGGTGCTG GGGCCATTTC CTTGCATCTG GGGGAGGGTC AGGGCTTTCC CTGTGGGAAC AAGTTAATAC 22960
ACAATGCACC TTACTTAGAC TTTACACGTA TTTAATGGTG TGCGACCCAA CATGGTCATT TGACCAGTAT 23030
TTTGGAAAGA ATTTAATTGG GGTGACCGGA AGGAGCAGAC AGACGTGGTG GTCCCCAAGA TGCTCCTTGT 23100
CACTACTGGG ACTGTTGTTC TGCCTGGGGG GCCTTGGAGG CCCCTCCTCC CTGGACAGGG TACCGTGCCT 23170
TTTCTACTCT GCTGGGCCTG CGGCCTGCGG TCAGGGCACC AGCTCCGGAG CACCCGCGGC CCCAGTGTCC 23240
ACGGAGTGCC AGGCTGTCAG CCACAGATGC CCAGGTCCAG GTGTGGCCGC TCCAGCCCCC GTGCCCCCAT 23310
GGGTGGTTTT GGGGGAAAAG GCCAAGGGCA GAGGTGTCAG GAGACTGGTG GGCTCATGAG AGCTGATTCT 23380
GCTCCTTGGC TGAGCTGCCC TGAGCAGCCT CTCCCGCCCT CTCCATCTGA AGGGATGTGG CTCTTTCTAC 23450
CTGGGGGTCC TGCCTGGGGC CAGCCTTGGG CTACCCCAGT GGCTGTACCA GAGGGACAGG CATCCTGTGT 23520
GGAGGGGCAT GGGTTCACGT GGCCCCAGAT GCAGCCTGGG ACCAGGCTCC CTGGTGCTGA TGGTGGGACA 23590
GTCACCCTGG GGGTTGACCG CCGGACTGGG CGTCCCCAGG GTTGACTATA GGACCAGGTG TCCAGGTGCC 23660
CTGCAAGTAG AGGGGCTCTC AGAGGCGTCT GGCTGGCATG GGTGGACGTG GCCCCGGGCA TGGCCTTCAG 23730
CGTGTGCTGC CGTGGGTGCC CTGAGCCCTC ACTGAGTCGG TGGGGGCTTG TGGCTTCCCG TGAGCTTCCC 23800
CCTAGTCTGT TGTCTGGCTG AGCAAGCCTC CTGAGGGGCT CTCTATTGCA GACAGCACTT GAAGAGGGTG 23870
CAGCTGCGGG AGCTGTCGGA AGCAGAGGTC AGGCAGCATC GGGAAGCCAG GCCCGCCCTG CTGACGTCCA 23940
GACTCCGCTT CATCCCCAAG CCTGACGGGC TGCGGCCGAT TGTGAACATG GACTACGTCG TGGGAGCCAG 24010
AACGTTCCGC AGAGAAAAGA GGGTGGCTGT GCTTTGGTTT AACTTCCTTT TTAAACAGAA GTGCGTTTGA 24080
GCCCCACATT TGGTATCAGC TTAGATGAAG GGCCCGGAGG AGGGGCCACG GGACACAGCC AGGGCCATGG 24150
CACGGCGCCA ACCCATTTGT GCGCACAGTG AGGTGGCCGA GGTGCCGGTG CCTCCAGAAA AGCAGCGTGG 24220
GGGTGTAGGG GGAGCTCCTG GGGCAGGGAC AGGCTCTGAG GACCACAAGA AGCAGCCGGG CCAGGGCCTG 24290
GATGCAGCAC GGCCCGAGGT CCTGGATCCG TGTCCTGCTG TGGTGCGCAG CCTCCGTGCG CTTCCGCTTA 24360
CGGGGCCCGG GGACCAGGCC ACGACTGCCA GGAGCCCACC GGGCTCTGAG GATCCTGGAC CTTGCCCCAC 24430
GGCTCCTGCA CCCCACCCCT GTGGCTGCGG TGGCTGCGGT GACCCCGTCA TCTGAGGAGA GTGTGGGGTG 24500
AGGTGGACAG AGGTGTGGCA TGAGGATCCC GTGTGCAACA CACATGCGGC CAGGAACCCG TTTCAAACAG 24570
GGTCTGAGGA AGCTGGGAGG GGTTCTAGGT CCCGGGTCTG GGTGGCTGGG GACACTGGGG AGGGGCTGCT 24640
TCTCCCCTGG GTCCCTATGG TGGGGTGGGC ACTTGGCCGG ATCCACTTTC CTGACTGTCT CCCATGCTGT 24710
CCCCGCCAGG CCGAGCGTCT CACCTCGAGG GTGAAGGCAC TGTTCAGCGT GCTCAACTAC GAGCGGGCGC 24780
GGCGCCCCGG CCTCCTGGGC GCCTCTGTGC TGGGCCTGGA CGATATCCAC AGGGCCTGGC GCACCTTCGT 24850
GCTGCGTGTG CGGGCCCAGG ACCCGCCGCC TGAGCTGTAC TTTGTCAAGG TGGGTGCCGG GGACCCCCGT 24920
GAGCAGCCCT GCTGGACCTT GGGAGTGGCT GCCTGATTGG CACCTCATGT TGGGTGGAGG AGGTACTCCT 24990
GGGTGGGCCG CAGGGAGTGC AGGTGACCCT GTCACTGTTG AGGACACACC TGGCACCTAG GGTGGAGGCC 25060
TTCAGCCTTT CCTGCAGCAC ATGGGGCCGA CTGTGCACCC TGACTGCCCG GGCTCCTATT CCCAAGGAGG 25130
GTCCCACTGG ATTCCAGTTT CCGTCAGAGA AGGAACCGCA ACGGCTCAGC CACCAGGCCC CGGTGCCTTG 25200
CACCCCAGTC CTGAGCCAGG GGTCTCCTGT CCTGAGGCTC AGAGAGGGGA CACAGCCCGC CCTGCCCTTG 25270
GGGTCTGGAG TGGTGGGGGT CAGAGAGAGA GTGGGGGACA CCGCCAGGCC AGGCCCTGAG GGCAGAGGTG 25340
ATGTCTGAGT TTCTGCGTGG CCACTGTCAG TCTCCTCGCC TCCACTCACA CAGGTGGATG TGACGGGCGC 25410
GTACGACACC ATCCCCCAGG ACAGGCTCAC GGAGGTCATC GCCAGCATCA TCAAACCCCA GAACACGTAC 25480
TGCGTGCGTC GGTATGCCGT GGTCCAGAAG GCCGCCCATG GGCACGTCCG CAAGGCCTTC AAGAGCCACG 25550
TAAGGTTCAC GTGTGATAGT CGTGTCCAGG ATGTGTGTCT CTGGGATATG AATGTGTCTA GAATGCAGTC 25620
GTGTCTGTGA TGCGTTTCTG TGGTGGAGGT ACTTCCATGA TTTACACATC TGTGATATGC GTGTGTGGCA 25690
CGTGTGTGTC GTGGTGCATG TATCTGTGGC GTGCATATTT GTGGTGTGTG TGTGTGTGGC ACGTGTGTGT 25760
CCATGGTGTG TGTGCCTGTG GTGTGCATGT GTGTGTGTCT GTGACACGTG CATGTTCATG CTGTGTGCTG 25830
CATGTCTGTG ATGTGCCTAT TTGTGGTGTG TGTGTGCATG TGTCCGTGAC ATATGCGTGT CTATGGCATG 25900
GGTGTGTGTG GCCCCTTGGC CTTACTCCTT CCTCCTCCAG GCATGGTCCG CACCATTGTC CTCACGCTCT 25970
CGGGTGCTGG TTTGGGGAGC TCCACATTCA GGGTCCTCAC TTCTAGCATG GGTGCCCCTG TCCTGTCACA 26040
GGGCTGGGCC TTGGAGACTG TAAGCCAGGT TTGAGAGGAG AGTAGGGATG CTGGTGGTAC CTTCCTGGAC 26110
CCCTGGCACC CCCAGGACCC CAGTCTGGCC TATGCCGGCT CCATGAGATA TAGGAAGGCT GATTCAGGCC 26180
TCGCTCCCCG GGACACACTC CTCCCAGAGC GGCCGGGGGC CTTGGGGCTC GGCAGGGGTG AAAGGGGCCC 26250
TGGGCTTGGG TTCCCACCCA GTGGTCATGA GCACGCTGGA GGGGTAAGCC CTCAAAGTCG TGCCAGGCCG 26320
GGGTGCAGAG GTGAAGAAGT ATCCCTGGAG CTTCGGTCTG GGGAGAGGCA CATGTGGAAA CCCACAAGGA 26390
CCTCTTTCTC TGACTTCTTG AGCT 26414

Contig 2:

```
TGTGGGATTG GTTTTCATGT GTGGGATAGG TGGGGATCTG TGGGATTGGT TTTTATGAGT GGGGTAACAC 70
AGAGTTCAAG GCGAGCTTTC TTCCTGTAGT GGGTCTGCAG GTGCTCCAAC AGCTTTATTG AGGAGACCAT 140
ATCTTCCTTT GAACTATGGT CGGGTTTATA GTAAGTCAGG GGTGTGGAGG CCTCCCCTGG GCTCCCTGTT 210
CTGTTTCTTC CACTCTGGGG TCGTGTGGTG CCTGCTGTGG TGTGTGGCCG GTGGGCAGGG CTTCCAGGCC 280
TCCTTGTGTT CATTGGCCTG GATGTGGCCC TGGCTACGCT CCGTCCTTGG AATTCCCCTG CGAGTTGGAG 350
GCTTTCTTTC TTTCTTTTTT TCTTTCTTTT TTTTTTTTT TGATAACAGA GTCTCGCTCT TTTTTGCCCA 420
GGCTGGAGTG GTTTGGCGTG ATCTTGGCTC ACTGCAACCT GTGCTTCCTG AGTTCAAGCA ATTCTCTTGC 490
CTCAGCCTCC CAAGTAGCTG GAATTATAGG CGCCCACCAC CATGCTGACT AATTTTTGTA ATTTTAGTAG 560
AGACGAGGTT TCTCCATGTT GGCCAGGCTG GTCTCGAACT CCTGACCTCA GGTGATCCTC CCACCTCGGC 630
CTCCCAAAGT GCTGGGATGA CAGGTGTGAA CCGCCGCGCC CGGCCGAGAC TCGCTTCCTG CAGCTTCCGT 700
GAGATCTGCA GCGATAGCTG CCTGCAGCCT TGGTGCTGAC AACCTCCGTT TTCCTTCTCC AGGTCTCGCT 770
AGGGGTCTTT CCATTTCATG ACTCTCTTCA CAGAAGAGTT TCACGTGTGC TGATTTCCCG GCTGTTTCCT 840
GCGTAATTGG TGTCTGCTGT TTATCGATGG CCTCCTTCCA TTTCCTTTAG GCTTTGTTTA TTGTTGTTTT 910
TCCGGCTCCT TGAAGGAAAA GTTTCGATTA TGGATGTTTG AACTTTCTTT TCTAAACAAG CATCTGAAGT 980
TGCCGTTTTC CCTCTAAAGC AGGGATCCCG AGGCCCCTGG CTGTGGAGTG GCACCGGTCT GGGGCCTGTT 1050
AGGAACCCGG CGCACAGCGG GAGGCTAGGT GGGGTGTGGG GAGCCAGCGT TCCCGCCTGA GCCCCGCCCC 1120
TCTCAGATCA GCAGTGGCAT GCGGTGCTCA GAGGCGCACA CACCCTACTG AGAACTGTGC GTGAGAGGGG 1190
TCTAGATTCT GTGCTCCTTA TGGGAATCTA ATGCCTGATG ATCTGAGGTG GAACCGTTTG CTCCCAAAAC 1260
CATCCCCTTC CCCACTGCTG TCCTGTGGAA AAATCGTCTT CCACGAAACC AGTCCCTGGT ACCACAATGG 1330
TTGGGGACCC TGTGCTAAAG ACCTGCTTCA GCAGCCTCTC GTCAGTGTTG ATATATTGGC TTTTCTGTGT 1400
TGAGTCCAGA ATAATTACGG ATTTCTGTGA TGCTTTCCGC CGACCTCAGA CCCATGGGCT ATTTGTGGGC 1470
GTGTTGCCTG CTCCTGGGTT GGGAAGGGTG CAGGCCCCAT GTACCTTCCT GTTACTGCCT TCCAGGTTGG 1540
TTCTCAGGGT TGAATCGTAC TCGATGTGGT TTTAGCCCAC GGCCCTGCCG CCAGCTCCTG GGGGCTGGGG 1610
AACATGCTGA AGCACAGAGT CACCGTGCGC GTCTTTTGAT GCCTCACAAG CTCGAGGCCT CCTGTGTCCG 1680
TGTTAGTGTG TGTCACGTGC CTGCTCACAT CCTGTCTTGG GGACGCAGGG GCTTAGCAGG TCCCGTAGTA 1750
AATGACAAGC GTCCTGGGGG AGTCTGCAGA ATAGGAGGTG GGGGTGCCGG TCTCTCTCCC GCGTCTTCAG 1820
ACTCTTCTCC TGCCTGTGCT GTGGCTGCAC CTGCATCCCT GCAATCCCTC CAGCACTGGG CTGGAGAGGC 1890
CCGGGAGCTC GAGTGCCACT TGTGCCACGT GACTGTGGAT GGCAGTCGGT CACGGGGGTC TGATGTGTGG 1960
TGACTGTGGA TGGCGGTTGG TCACAGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTCG TGGGGTCTGA 2030
TGTGGTGACT GTGGATGGCG GTCGTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTCG TGGGGTCTGA 2100
TGTGGTGACT GTGGATGGCG GTCGTGGGGT CTGATGTGGT GACTGTGGAT GGCGGTCGTG GGGTCTGATG 2170
TGGTGACTGT GGATGGCAGT CGTGGGGTCT GATGTGTGGT GACTGTGGAT GGCGGTCGTG GGGTCTGATG 2240
TGGTGACTGT GGATGGCAGT CGTGGGGTCT GATGTGTGGT GACTGTGGAT GGCGGTCGTG GGGTCTGATG 2310
TGTGGTGACT GTGGATGGCG GTCGTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTCG TGGGGTCTGA 2380
TGTGTGGTGA CTGTGGATGG CGGTCGTGGG GTCTGATGTG GTGACTGTGG ATGGCGGTCG TGGGGTCTGA 2450
TGTGTGGTGA CTGTGGATGG TGATCGGTCA CAGGGGTCTG ATGTGTGGTG ACTGTGGATG GCGGTCGTGG 2520
GGTCTGATGT GTGGTGACTG TGGATGGTGA TCGGTCACAG GGGTCTGATG TGTGGTGACT GTGGATGGCG 2590
GTCGTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTTG GTCCCGGGGG TCTGATGTGT GGTGACTGTG 2660
GATGGCGATC GGTCACAGGG GTCTGATGTG TGGTGACTGT GGATGGCGGT CGTGGGGTCT GATGTGTGGT 2730
GACTGTGGAT GGCGGTCGTG GGGTCTGATG TGTGGTGACT GTGGATGGCG GTCGTGGGGT CTGATGTGGT 2800
GACTGTGGAT GGCGGTCGTG GGGTCTGATG TGGTGACTGT GGATGGCGGT CGTGGGGTCT GATGTGTGGT 2870
GACTGTGGAT GGCGGTTGGT CCCGGGGGTC TGATGTGTGG TGACTGTGGA TGGCGGTCGT GGGGTCTGAT 2940
GTGGTGACTG TGGATGGCAG TCGTGGGGTC TGATGTGTGG TGACTGTGGA TGGCGGTCGT GGGGTCTGAT 3010
GTGTGGTGAC TGTGGATGGC GGTCGTGGGG TCTGATGTGT GGTGACTGTG GATGGCGGTC GTGGGGTCTG 3080
ATGTGTGGTG ACTGTGGATG GCGGTCGTGG GGTCTGATGT GGTGACTGTG GATGGCGGTC GTGGGGTCTG 3150
ATGTGTGGTG ACTGTGGATG GTGATCGGTC ACAGGGGTCT GATGTGTGGT GACTGTGGAT GGCGGTCGTG 3220
GGGTCTGATG TGTGGTGACT GTGGATGGCG GTCGTGGGGT CTGATGTGGT GACTGTGGAT GGCGGTCGTG 3290
GGGTCTGATG TGTGGTGACT GTGGATGGCG GTCGTAGGGT CTGATGTGTG GTGACTGTGG ATGGCAGTCG 3360
GTCACAGGGG TCTGATGTGT GGTGACTGTG GATGGCGGTC GTGGGGTCTG ATGTGTGGTG ACTGTGGATG 3430
GCGGTCGTGG GGTCTGATGT GTGGTGACTG TGGATGGCGG TCGTGGGGTC TGATGTGTGG TGACTGTGGA 3500
TGGCGGTCGT GGGGTCTGAT GTGGTGACTG TGGATGGTGA TCGGTCACAG GGGTCTGATG TGTGGTAGCT 3570
GCAGGTGGAG TCCCAGGTGT GTCTGTAGCT ACTTTGCGTC CTCGGCCCCC CGGCCCCCGT TTCCCAAACA 3640
GAAGCTTCCC AGGCGCTCTC TGGGCTTCAT CCCGCCATCG GGCTTGGCCG CAGGTCCACA CGTCCTGATC 3710
GGAAGAAACA AGTGCCCAGC TCTGGCCGGG GCAGGCCACA TTTGTGGCTC ATGCCCTCTC CTCTGCCGGC 3780
AGGTCTCTAC CTTGACAGAC CTCCAGCCGT ACATGCGACA GTTCGTGGCT CACCTGCAGG AGACCAGCCC 3850
GCTGAGGGAT GCCGTCGTCA TCGAGCAGGT CTGGGCACTG CCCTGCAGGG TTGGGCACGG ACTCCCAGCA 3920
GTGGGTCCTC CCCTGGGCAA TCACTGGGCT CATGACCGGA CAGACTGTTG GCCCTGGGGG GCAGTGGGGG 3990
GAATGAGCTG TGATGGGGGC ATGATGAGCT GTGTGCCTTG GCGAAATCTG AGCTGGGCCA TGCCAGGCTG 4060
CGACAGCTGC TGCATTCAGG CACCTGCTCA CGTTTGACTG CGCGGCCTCT CTCCAGTTCC GCAGTGCCTT 4130
TGTTCATGAT TTGCTAAATG TCTTCTCTGC CAGTTTTGAT CTTGAGGCCA AAGGAAAGGT GTCCCCCTCC 4200
TTTAGGAGGG CAGGCCATGT TTGAGCCGTG TCCTGCCCAG CTGGCCCCTC AGTGCTGGGT CTGAGGCCAA 4270
AGGAAACGTG TCCCCCTTCT TAGGAGGACG GGCCGTGTTT GAGCCACGCC CCGCTGAGCG GGCCTCTCAG 4340
TGCTGGGTCT GTCCACGTGG CCCTGTGGCC CTTTGCAGAT GTGGTCTGTC CACGTGGCCC TGTGGCTCTT 4410
TGCAGATGCC TGTTAGCACT TGCTCGGCTC TAGGGGACAG TCGTGTCCAC CGCATGAGGC TCAGAGACCT 4480
CTGGGCGAAT TTCCTTGGCT CCCAGGGTGG GGGTGGAGGT GGCCTGGGCT GCTGGGACCC AGACCCTGTG 4550
CCCGGCAGCT GGGCAGCAAC TCCTGGATCA CATATGCCAT CCGGGCCACG GTGGGCTGTG TGGGTGTGAG 4620
CCCAGCTGGA CCCACAGGTG GCCCAGAGGA GACGTTCTGT GTCACACACT CTGCCTAAGC CCATGTGTGT 4690
CTGCAGAGAC TCGGCCCGGC CAGCCCACGA TGGCCCTGCA TTCCAGCCCA GCCCCGCACT TCATCACAAA 4760
CACTGACCCC AAAAGGGACG GAGGGTCTTG GCCACGTGGT CCTGCCTGTC TCAGCACCCA CCGGCTCACT 4830
CCCATGTGTC TCCCGTCTGC TTTCGCAGAG CTCCTCCCTG AATGAGGCCA GCAGTGGCCT CTTCGACGTC 4900
TTCCTACGCT TCATGTGCCA CCACGCCGTG CGCATCAGGG GCAAGTGAGT CAGGTGGCCA GGTGCCATTG 4970
CCCTGCGGGT GGCTGGGCGG GCTGGCAGGG CTTCTGCTCA CCTCTCTCCT GCCCCTTCCC CACTGNCCTT 5040
```

```
CTGCCCGGGG CCACCAGAGT CTCCTTTTCT GGCCCCCGCC CCCTCCGGCT CCTGGGCTGC AGGCTCCCGA 5110
GGCCCCGGAA ACATGGCTCG GCTTGCGGCA GCCGGAGCGG AGCAGGTGCC ACACGAGGCC TGGAAATGGC 5180
AAGCGGGGTG TGGAGTTGCT CCTGCGTGGA GGACGAGGGG CGGGGGGTGT GTCTGGGTCA GGTGTGCGCC 5250
GAGCGTTTGA GCCTGCAGCT TGTCAGCTCC AAGTTACTAC TGACGCTGGA CACCCGGCTC TCACACGCTT 5320
GTATCTCTCT CTCCCGATAC AAAAGGATTT TATCCGATTC TCATTCCTGT CCCTGTCGTG TGACCCCCGC 5390
GAGGGCGCGG GCTCTTCTCT CTGTGACTAG ATTTCCCATC TGGAAAGTGC GGGGTTGACC GTGTAGTTTG 5460
CTCCTCTCGG GGGGCCTGTG GTGGCCATGG GGCAGGCGGC CTGGGAGAGC TGCCGTCACA CAGCCACTGG 5530
GTGAGCCACA CTCACGGTGG TAGAGCCACA GTGCCTGGTG CCACATCACG TCCTCTGGAT TTTAAGTAAA 5600
ACCACACACC TCCCGGCAGG CATCTGCCTG CGACCCTGTG TGTGCCTGGG GAGAGTGGTA GCACGGAGGA 5670
AATTCGTGCA CACTCAAGGT CATCAGCAAG GTCATCCGCA GTCAGGTGGA ACGTGGAGGC CTCTCTCTGG 5740
GATCGTCTCC AGCGGATAAA GGACTGTGCA CAGCTTCGGA AGCTTTTATT TAAAAATATA ACTATTAATT 5810
ATTGCATTAT AAGTAATCAC TAATGGTATC AGCAATTATA ATATTTATTA AAGTATAATT AGAAATATTA 5880
AGTAGTACAC ACGTTCTGGA AAAACACAAA TTGCACATGG CAGCAGAGTG AATTTTGGCC GAGGGACACG 5950
TGTGCACATG TGTGTAAGCG GCCCCCAGGC CCACAGAATT CGCTGACAAA GTCACCTCCC CAGAGAAGCC 6020
ACCACGGGCC TCCTTCGTGG TCGTGAATTT TATTAAGATG GATCAAGTCA CGTACCGTCC ACGTGTGGCA 6090
GGGCTTTGGG GAATGTGAGG TGATGACTGC GTCCTCATGC CCTGACAGAC AGGAGGTGAC TGTGTCTGTC 6160
CTGTCCCTAG GACACGGACA GGCCCGAAGC TCTAGTCCCC ATCGTGGTCC AGTTTGGCCT CTGAATAAAA 6230
ACGTCTTCAA AACCTGTTGC CCCAAAAACT AAGAACAGAG AGAGTTTCCC ATCCCATGTG CTCACAGGGG 6300
CGTATCTGCT TGCGTTGACT CGCTGGGCTG GCCGGACTCC TAGAGTTGGT GCGTGTGCTT CTGTGCAAAA 6370
AGTGCAGTCC TCTTGCCCAT CACTGTGATA TCTGCACCAG CAAGGAAAGC CTCTTTTCTT TTCTTTCTTT 6440
TTTTTTTTTT GAGACGGAAC GTCACTGTTG TCTGCCTGGG CTTGAGTGCA GTGGCGCGAT CTCAACTCAC 6510
TGCAACCTCC GCCTCCCGGG TTCCAGCATT TCTCCTGCCT CAGCCTCCCG AGCAGCTGAG ATTACAGGCA 6580
CCCACCCCCT GCGCCTGGCT AATTTTTGTA TTTTTAGTAG AGAGGGGTTT TTGCCATGTT GGCCAGGCTG 6650
GTCTCGAACT CCTGACCTCA GGTGATCCAC CCACCTCGGC CTCCCAAAGT GCTGGGATTA CAGGTGTGAG 6720
CCATCACGCC CAGCCGGAAA GCCTCTTTTT AAGGTGACCA CCTATAGCGC TTCCCGAAAA TAACAGGTCT 6790
TGTTTTTGCA GTAGGCTGCA AGCGTCTCTT AGCAACAGGA GTGGCGTCCT GTGGGCTCTG GGGATGGCTG 6860
AGGGTCGCGT GGCAGCCATG CCTTCTGTGT GCACCTTTAG GTTCCACGGG GCTATTCTGC TCTCACTGTT 6930
TGTCTGAAAA CGCACCCTTG GCATCCTTGT TTGGAGAGTT TCTGCTTCTC GTTGGTCATG CTGAAACTAG 7000
GGGCAAGGTT GTATCCGTTG GCGCGCAGCG GCTACATGTA GGGTCATGAG TCTTTCACCG TGGACAAATT 7070
CCTTGAAAAA AAAAAAAGGA GTCCGGTTAA GCATTCATTC CGGGTCAAGT GTCTGGTTCT GTGAATAAAC 7140
TCTAAGATTT AAGAAACCTT AATGAAAGAA AACCTTGATG ATTCAGAGCA AGGATGTGGT CACACCTGTG 7210
GCTGGATCTG TTTCAGCCGC CCCAGTGCAT GGTGAGAGTG GGGAGCAGGG ATTGTTTGTT CAGAGGTCTC 7280
ATCTGGTATG TTTCTGAGGT GTTTGCCGGC TGAATGGTAG ACGTGTCGTT TGTGTGTATG AGGTTCTGTG 7350
TCTGTGTGTG GCTCGGTTTG AGTGTACGCA TGTCCAGCAC ATGCCCTGCC CGTCTCTCAC CTGTGTCTTC 7420
CCGCCCCAGG TCCTACGTCC AGTGCCAGGG GATCCCGCAG GGCTCCATCC TCTCCACGCT GCTCTGCAGC 7490
CTGTGCTACG GCGACATGGA GAACAAGCTG TTTGCGGGGA TTCGGCGGGA CGGGTGAGGC CTCCTCTTCC 7560
CCAGGGGGGC TTGGGTGGGG GTTGATTTGC TTTTGATGCA TTCAGTGTTA ATATTCCTGG TGCTCTGGAG 7630
ACCATGACTG CTCTGTCTTG AGGAACCAGA CAAGGTTGCA GCCCCTTCTT GGTATGAAGC CGCACGGGAG 7700
GGGTTGCACA GCCTGAGGAC TGCGGGCTCC ACGCAGGCTC TGTCCAGCGG CCATGTCCAG AGGCCTCAGG 7770
GCTCAGCAGG CGGGAGGGCC GCTGCCCTGC ATGATGAGCA TGTGAATTCA ACACCGAGGA AGCACACCAG 7840
CTTCTGTCAC GTCACCCAGG TTCCGTTAGG GTCCTTGGGG AGATGGGGCT GGTGCAGCCT GAGGCCCCAC 7910
ATCTCCCAGC AGGCCCTCGA CAGGTGGCCT GGACTGGGCG CCTCTTCAGC CCATTGCCCA TCCCACTTGC 7980
ATGGGGTCTA CACCCAAGGA CGCACACACC TAAATATCGT GCCAACCTAA TGTGGTTCAA CTCAGCTGGC 8050
TTTTATTGAC AGCAGTTACT TTTTTTTTTT TAATACTTTA AGTTCTAGGG TACATGTGCA CGACGTGCAG 8120
GTTAGTTACA TATGTATACA TGTGCCATGT TGGTGTGCTG CACCCATTAA CTCATCATTT ACATTAGGTA 8190
TATCTCCTAA TGCTATCCCT CCCCACTCCC CCCATCCCAT GACAGGCCCT GGTGTGTGAT GTTCCCCACC 8260
CTGTGTCCAA GTGTTCTCAT TGTTCAGTTC CCACCTGTGA GTGAGAACAT GTGGTGTTTG GTTTTCTTTC 8330
CTTGCAATAG TTTGCTCAGA GTGATGGTTT CCAGCTTCGT CCATGTCCCT ACAAAGGACA TGAACTCATC 8400
CTTTTTTATG ACTGCATAGT ATTCCGTGGT GTATATGTGC CACATTTTCT TAATCCAGTC TATCATCGAT 8470
GGACATTTGG GTTGGTTGCA AGTCTTTGCT ACTGTGAATA GTGCCGCAAT AAACATACGT GTGCATGTGT 8540
CTTTATAGCA GCATGATTTA TAATCCTTTG GGTATATACC CAGTAATGGG ATGGCTGGGT CAAATGGTAT 8610
TTCTAGTTCT AGATCCTTGA GGAATCACCA CACTGTCTTC CACAATGGTT GAACTAGTTT ACACTCCCAC 8680
CAACAGTGTA AAAGTGTTCT GGTGCTGGAG AGGATGTGGA CAGCAGTTAT TTTTTTATGA AAATAGTATC 8750
ACTGAACAAG CAGACAGTTA GTGAAGGATG CGTCAGGAAG CCTGCAGGCC ACACAGCCAT TTCTCTCGAA 8820
GACTCCGGGT TTTTCCTGTG CATCTTTTGA AACTCTAGCT CCAATTATAG CATGTACAGT GGATCAAGGT 8890
TCTTCTTCAT TAAGGTTCAA GTTCTAGATT GAAATAAGTT TATGTAACAG AAACAAAAAT TTCTTGTACA 8960
CACAACTTGC TCTGGGATTT GGAGGAAAGT GTCCTCGAGC TGGCGGCACA CTGGTCAGCC CTCTGGGACA 9030
GGATACCTCT GGCCCATGGT CATGGGGCGC TGGGCTTGGG CCTGAGGGTC ACACAGTGCA CCATGCCCAG 9100
CTTCCTGTGG ATAGGATCTG GGTCTCGGAT CATGCTGAGG ACCACAGCTG CCATGCTGGT AAAGGGCACC 9170
ACGTGGCTCA GAGGGGGCGA GGTTCCCAGC CCCAGCTTTC TTACCGTCTT CAGTTATTTT TCCCTAAGAG 9240
TCTGAGAAGT GGGGCCGCGC CTGATGGCCT TCGTTCGTCT TCAGCTGGCA CAGAATTGCA CAAGCTGATG 9310
GTAAACACTG AGTACTTATA ATGAATGAGG AATTGCTGTA GCAGTTAACT GTAGAGAGCT CGTCTGTTGG 9380
AAAGAAATTT AAGTTTTTCA TTTAACCGCT TTGGAGAATG TTACTTTATT TATGGCTGTG TAAATTGTTT 9450
GACATTCAGT CCCTCGTAGA CAGATACTAC GTAAAAAGTG TAAAGTTAAC CTTGCTGTGT ATTTTCCCTT 9520
ATTTTAGGCT GCTCCTGCGT TTGGTGGATG ATTTCTTGTT GGTGACACCT CACCTCACCC ACGCGAAAAC 9590
CTTCCTCAGG TGAGGCCCGT GCCGTGTGTC TGTGGGGACC TCCACAGCCT GTGGGCTTTG CAGTTGAGCC 9660
CCCCGTGTCC TGCCCCTGGC ACCGCAGCGT TGTCTCTGCC AAGTCCTCTC TCTCTGCCGG TGCTGGATCC 9730
GCAAGAGCAG AGGCGCTTGG CCGTGCACCC AGGCCTGGGG GCGCAGGGGC ACCTTCGGGA GGGAGTGGGT 9800
ACCGTGCAGG CCCTGGTCCT GCAGAGACGC ACCCAGGTTA CACACGTGGT GAGTGCAGGC GGTGACCTGG 9870
CTCCTGCTGC TCTTTGGAAA GTCAAGAGTG GCGGCTCCTG GGGCCCCAGT GAGACCCCCA GGAGCTGTGC 9940
ACAGGGCCTG CAGGGCCGAG GCGGCAGCCT CCTCCCCAGG GTGCACCTGA GCCTGCGGAG AGCAGGAGCT 10010
GCTGAGTGAG CTGGCCCACA GCGTTCGCTG CGGTCACGTT CCTGCGTGGG GTTGTTTGGG ATCGGTGGGA 10080
GAATTTGGAT TTGCTGAGTG CTGCTGTCTT GAACCACGGA GATGGCTAGG AGTGGGTTTC AGAGTTGATT 10150
TTTGTGAATC AAACTAAAAT CAGGCACAGG GGACCTGGCC TCAGCACAGG GGATTGTCCA ATGTGGTCCC 10220
CCTCAAGGGC GCCCCACAGA GCCGGTGGGC TTGTTTTAAA GTGCGATTTG ACGAGGGACG AGAAACCTTG 10290
AAAGCTGTAA AGGGAACCCT CAGAAAATGT GGCCGCCAGG GGTGGTTTCA GGTGCTTTGC TGGGCTGTGT 10360
TTGTGAAAAC CCATTTGGAC CCGCCCTCCA AGTCCACCCT CCAGGTCCAC CCTCCAGGGC CGCCCTGGGC 10430
```

```
TGGGGGTATG CCTGGCGTTC CTTGTGCCGC AGCCCGGAGC ACAGCAGGCT GTGCACATTT AAATCCACTA 10500
AGATTCACTC GGGGGGAGCC CAGGTCCCAA GCAACTGAGG GCTCAGGAGT CCTGAGGCTG CTGAGGGGAC 10570
AGAGCAGACG GGGAACGCTG CTTCTGTGTG GCAAGTTCCT GAGGGTGCTG GCCAGGGAGG TGGCTCAGAG 10640
TGTATGTTGG GGTCCCACCG GGGGCAGAAC TCTGTCTCTG ATGAGTCGGC AGCCATGTAA CAGGAAGGGG 10710
TGGCCACAGG GAGCTGGGAA TGCACCAGGG GAGCTGCGCA GCTGGCCGAG GTCCCAGGGC CAGGCCACAG 10780
GAAGGGCAGG GGGACGCCCG GGGCCACAGC AGAGGCCGCA GGAAGGGAAG GGGATGCCCA GGCCAGAGCA 10850
GAGGCTACCG GGCACAGGGG GGCTCCCTGA GCTGGGTGAG CGAGGCTCAT GACTCGGCGA GGGAACCTCC 10920
TTGACGTGAA GCTGACGACT GGTGTTGCCC AGCTCACAGC CCAGCCAGGT CCCGCGCCTG AGCAGGAACT 10990
CAGAACCCTC CCCTTTGTCT AAAGCACAGC AGATGCCTTC AGGGCATCTA GGAGAAAACA GGCAAAGTCG 11060
TTGAGAAACG TCTTAAAAGA AGGTGGGATG GTGGCAATTT CTTGTCCAGA TTTTAGTCTG CCCCGGACCA 11130
CAGATGAGTC TATAACGGGA TTGTGGTGTT GCCATGGGGA CACATGAGAT GGACCATCAC AGAGGCCACT 11200
GGGGCTGCAC CTCCCATCTG AGTCCTGGCT GTCCCGGGTC CAGGCCAGGT TCTTGCATGC TCACCTACCT 11270
GTCCTGCCCG GGAGACAGGG AAAGCACCCC GAAGTCTGGA GCAGGGCTGG GTCCAGGCTC CTCAGAGCTC 11340
CTGCCAGGCC CAGCACCCTG CTCCAAATCA CCACTTCTCT GGGGTTTTCC AAAGCATTTA ACAAGGGTGT 11410
CAGGTTACCT CCTGGGTGAC GGCCCCGCAT CCTGGGGCTG ACATTGCCCC TCTGCCTTAG GACCCTGGTC 11480
CGAGGTGTCC CTGAGTATGG CTGCGTGGTG AACTTGCGGA AGACAGTGGT GAACTTCCCT GTAGAAGACG 11550
AGGCCCTGGG TGGCACGGCT TTTGTTCAGA TGCCGGCCCA CGGCCTATTC CCCTGGTGCG GCCTGCTGCT 11620
GGATACCCGG ACCCTGGAGG TGCAGAGCGA CTACTCCAGG TGAGCGCACC TGGCCGGAAG TGGAGCCTGT 11690
GCCCGGCTGG GGCAGGTGCT GCTGCAGGGC CGTTGCGTCC ACCTCTGCTT CCGTGTGGGG CAGGCGACTG 11760
CCAATCCCAA AGGGTCAGAG GCCACAGGGT GCCCCTCGTC CCATCTGGGG CTGAGCAGAA ATGCATCTTT 11830
CTGTGGGAGT GAGGGTGCTC ACAACGGGAG CAGTTTTCTG TGCTATTTTG GTAAAAGGAA ATGGTGCACC 11900
AGACCTGGGT GCACTGAGGT GTCTTCAGAA AGCAGTCTGG ATCCGAACCC AAGACGCCCG GGCCCTGCTG 11970
GGCGTGAGTC TCTCAAACCC GAACACAGGG GCCCTGCTGG GCATGAGTCC CTCTGAACCC GAGACCCTGG 12040
GGCCCTGCTG GGCGTGAGTC TCTCCGAACC CAGAGACTTC AGGGCCCTTT TGGGCGTGAG TCTCTCCGCT 12110
GTGAGCCCCA CACTCCAAGG CTCATCCACA GTCTACAGGA TGCCATGAGT TCATGATCAC GTGTGACCCA 12180
TCAGGGGACA GGGCCATGGT GTGGGGGGGG TCTCTACAAA ATTCTGGGGT CTTGTTTCCC CAGAGCCCGA 12250
GAGCTCAAGG CCCCGTCTCA GGCTCAGACA CAAATGAATT GAAGATGGAC ACAGATGCAG AAATCTGTGC 12320
TGTTTCTTTT ATGAATAAAA AGTATCAACA TTCCAGGCAG GGCAAGGTGG CTCACACCTA TAATCCCAGC 12390
ACTTTGGGAG GCCGAGGTGG GTGGATCACT TGAGGCCAGG AGTTTGAGGC CAACCTAACC AACATAGTGA 12460
AATTCCATTT CTACTTAAAA AATACAAAAA TTAGCCTGGC CTGGTGGCAC ACGCCTGTAG TCCCCGCTAT 12530
GCGGGAGGCT GAGGCAGGAG AATCATTTGA ACCCAGGAGG CAGAGGTTGC AGTGAGCCGA GATCACACCA 12600
CTGCACTCCA GCCTGGGCAA CAGAGTGAGA CTTCATCTTA AAAAAAAAAA AAAAAGTATC AGCATTCCAA 12670
AACCATAGTG GACAGGTGTT TTTTTATTCT GTCCTTCGAT AATATTTACT GGTGCTGTGC TAGAGGCCGG 12740
AACTGGGGGT GCCTTCCTCT GAAAGGCACA CCTTCATGGG AAGAGAAATA AGTGGTGAAT GGTTGTTAAA 12810
CCAGAGGTTT AAACTGGGGT CCTGTCGTTC TGAGTTAACA GTCCAGATCT GGACTTTGCC TCTTTCCAGA 12880
ATGCTCCCTG GGGTTTGCTT CATGGGGGAG CAGCAGGTGT GGACACCCTC GTGATGGGGG AGCAGCAGGT 12950
GCAGACGCCC TCATGATGGG GGAGTGGCAG GTGCAGACAC CCTTGTGCAT GGTGCCCAGC ATGTCCCTGT 13020
TGCAGCTCCC TCCCCACAAG GATGCCGGTC TCCTGTGCTC CCCACAGTCC CTGCTTCCCT CTCACAGCCT 13090
TACCTGGTCC TGGCCTCCAC TGGCTTTGTC TGCATGATTT CCACATTTCC TGGGCTCCCA GCACCTCTTC 13160
GCCTCTCCCA GGCACCTCTG CAGTGCTGGC CATACCAGTC AGCTGTGAAC TGTCCACTGC TTATTTTGCT 13230
CCCCATGAAA TGTATTTTTT AGGACAGGCA CCCCTGGTTC CAGCCTCTGG CACAGCATCA GTGAATGTTA 13300
TTGAAGGACA AAGGACAGAC AAACAAATCA GGAAAATGGG TTCTCTCTAA ACACATTGCA AAGCCACAGA 13370
GGCTAGTGCA GGATGGGTGG GCATCAGGTC ATCAGATGTG GGTCCAATGC CAGAATATTC TGTGCTCCCA 13440
AAGGCCACTT GGTCAGAGTG TGTGCTTGCA GAGGTGGCTC TAAAAGCTCA GCAGTGGAGG CAGTGGTTCG 13510
CCATACTCAG GGTGAACTCA CATCCTCTGT GTCTGAAGTA TACAGCAGAG GCTTGAAGGG CATCTGGGAG 13580
AAGAAAACAG GCAAAATGAT TAAGAAAAGT GAAAAAGGAA AAGTGGTAAG ATGGGAATTT TCTTGTCCAG 13650
ATTTTAGTCT CCCAAACCAC AGCTCAGATG GTAGAATGTG GTCAGAACTG ATGGACAGAA CAATAGAACA 13720
AAACGGAAGC CCTATCTCTC AGAAACGTGT GTTAATGTGG TATGTGGCAC AGCTGATGGA AAAGAGAGTG 13790
TGTGTGTAAT TTTTTTTTCT GAGAAAACTG ACTGGAAGCA AATAAGTTGT GTCTTTACAG CATATACCAG 13860
AGCAGATTCT AGGTAGAAGA GGAGACACAT GCAAACAACA CCAGCAACAG AAATAAAACA AAAGACTCAA 13930
AGGGAAGGGA GGTGAACGTT CCCTGGTTTG GTGTTGGGGA AGGACACACA GGGAGGCGGA TGAAACCAGT 14000
GAGGCAACGG GCATTGCTTT CACTGCAGAG AAACTCAGCT TGCCTGAGCC ACAGTGAAAA TGGCCATTCC 14070
CTGGAGCGTT TGTGCACGTG ATTTATTTAA GGCGCCCTGT GAGGTCCTGC ACATTCATCC TCTCACTTTG 14140
TTCTCCTAAC CACCTGAGAG GTAGAGGAGG AAAGGCTCCA GGGGAGCAGC CGCCCTTGGT CACCCAGCTG 14210
GCAAAGGGCA TGCATGATTG CAGCCTGGCC TCCTGCTCCG GGGCCCTTGC TCTGCCCGAG GACCCCACAC 14280
AAGTCAGACC CATAGGCTCA GGGTGAGCCG GAGCCCAAGG TCGTGTTGGG GATGGCTGTG AAAGAAGAAA 14350
TGGACGTCTG ATGCACACTT GGGAAGGTCC TACCAGCAGC GTCAAAGAAA TGCATGTGAA ACTGACAGCG 14420
AGACCCATCC CTCAAAGAAA CGCACGTGAA ACTGATGGCG AGACCTGTCC CCATCCCTCA TGCTGGCTCC 14490
TTTTCTGGGC TTGCCAAGAG CCAGCATCAG GTTGAGGCAA GCTGGAAAGA CTTTTCTGGA AAGCAGCTTG 14560
TTTGCATGGA AGTCCTCACA ATGTCCTGTG TCTTCCCAGT AATTCCACTT CTGAAGTGAC CAGACATTAT 14630
CACGGGTCTT ATTTACCATT TCCAGTGTTC CAGGCAGGGG GACTTGCCAC AGCAAGTCAC GAACCTGCCC 14700
AAATACAGGG CTAAGGAGAT ATTATGCATC ACAAAACTTG CTCTGCCATT AAACATTTTT CAAAGAATTT 14770
TTGAAGAATG TTTAATGGCA CAAAACGTTT ATTTCAATGT AGCAGTGTTC AAAGCTGGAT GTAAAAGAAC 14840
ACACCCCAGG AGCCTGCCGT GAATGTCATG TGTGTTCATC TTTGGACATG GACATACATG GGCAGTGAGT 14910
GGTGGTGAGG CCCTGGAGGA CATCGGTGGG ATGCCTCCAT CCTGCCCCTC TGGAGACACC ATGTGTGCCA 14980
CGTGCACTCA CTGGAGCCCT GTTTAGCTGG TGCCACCTGG CTCTTCCATC CCTGAGATTC AAACACAGTG 15050
AGATTCCCCA CGCCCAACTC AGTGTTCTCC CACAAAAAAC CTGAGTCACA CCTGTGTTCA CTCGAGGGAC 15120
GCCCGGGAGC CAGGGCTCCA CAGTTTATTA TGTGTTTTTG GCTGAGTTAT GTGCAGATCT CATCAGGGCA 15190
GATGATGAGT GCACAAACAC GGCCGTGCGA GGTTTGGATA CACTCAACAT CACTAGCCAG GTCCTGGTGG 15260
AGTTTGGTCA TGCAGAGTCT GGATGGCATG TAGCATTTGG AGTCCATGGA GTGAGCACCC AGCCCCCTCG 15330
GGCTGCAGCG CATGCCCCAG GCAGGACAAG GAAGCGGGAG GAAGGCAGGA GGCTCTTTGG AGCAAGCTTT 15400
GCAGGAGGGG GCTGGGTGTG GGGCAGGCAC CTGTGTCTGA CATTCCCCCC TGTGTCTCAG CTATGCCCGG 15470
ACCTCCATCA GAGCCAGTCT CACCTTCAAC CGCGGCTTCA AGGCTGGGAG GAACATGCGT CGCAAACTCT 15540
TTGGGGTCTT GCGGCTGAAG TGTCACAGCC TGTTTCTGGA TTTGCAGGTG AGCAGGCTGA TGGTCAGCAC 15610
AGAGTTCAGA GTTCAGGAGG TGTGTGCGCA AGTATGTGTG TGTGTGTGTG CGCGCGTGCC TGCAAGGCTG 15680
ATGGTGACTG GCTGCACGTA AGAGTGCACA TGTACGCATA TACACGTGAG CACATACATG TGTGCATGTG 15750
TGTACATGAA GGCATGGCAG TGTGTGCACA GGTGTGCAAG GGCACAAGTG TGTGCACATG CGAATGCACA 15820
```

```
CCTGACATGC ATGTGTGTTC GTGCACAGTC GTGTGGGCAT TCACGTGAGG TGCATGCGTG TGGGTGTGCA 15890
GTGTGAGTAG CATGTGTGCA CATAACATGT ATTGAGGGGT CCTCGTGTTC ACCCCGCTAG GTCCTCAGCA 15960
CCAGTGCCAC TCCTTACAGG ATGAGACGGG GTCCCAGGCC TTGGTGGGCT GAGGCTCTGA AGCTGCAGCC 16030
CTGAGGGCAT TGTCCCATCT GGGCATCCGC GTCCACTCCC TCTCCTGTGG GCTTCTGTGT CCACTCCCCC 16100
TCTCCTGTGG GCATTTACAT CCACTCCACT CCCTCTCTCC TGTGGGCATC CGCGTCCACT CCCCCTCTCT 16170
GTGGGCATCT GCGTCCACCT CCCCTCTCTG TGGGCATTTG CGTCCACTCC CTCTCCTGGT TCCTTCCTGT 16240
CTTGGCCGAG CCTCGGGGGC AGGCAGATGA CACAGAGTCT TGACTCGCCC AGGGTGGTTC GCAGCTGCCG 16310
GGTGAGGGCC AGGCCGGATT TCACTGGGAA GAGGGATAGT TTCTTGTCAA AATGTTCCTC TTTCTTGTTC 16380
CATCTGAATG GATGATAAAG CAAAAAGTAA AAACTTAAAA TCCCAGAGAG GTTTCTACCG TTTCTCACTC 16450
TTTCTTGGCG ACTCTAGGTG AACAGCCTCC AGACGGTGTG CACCAACATC TACAAGATCC TCCTGCTGCA 16520
GGCGTACAGG TGAGCCGCCA CCAAGGGGTG CAGGCCCAGC CTCCAGGGAC CCTCCGCGCT CTGCTCACCT 16590
CTGACCCGGG GCTTCACCTT GGAACTCCTG GGTTTTAGGG GCAAGGAATG TCTTACGTTT TCAGTGGTGC 16660
TGCTGCCTGT GCACAGTTCT GTTCGCGTGG CTCTGTGCAA AGCACCTGTT CTCCATCTCT GGGTAGTGGT 16730
AGGAGCCGGT GTGGCCCCAG GTGTCCCCAC TGTGCCTGTG CACTGGCCGT GGGACGTCAT GGAGGCCATC 16800
CCAGGGCAGC AGGGGCATGG GGTAAAGAGA TGTTTATGGG GAGTCTTAGC AGAGGAGGCT GGGAAGGTGT 16870
CTGAACAGTA GATGGGAGAT CAGATGCCCG GAGGATTTGG GGTCTCAGCA AAGAGGGCCG AGGTGGGTGC 16940
AGGTGAGGGT CGCTGGCCCC ACCCCCGGGA AGGTGCAGCA GAGCTGTGGC TCCCCACACA GCCCGGCCAG 17010
CACCTGTGCT CTGGGCATGG CTGTGCTCCT GGAACGTTCC CTGTCCTGGC TGGTCAGGGG GTGCCCCTGC 17080
CAAGAATCGA CAACTTTATC ACAGAGGGAA GGGCCAATCT GTGGAGGCCA CAGGGCCAGC TTCTGCCTGG 17150
AGTCAGGGCA GGTGGTGGCA CAAGCCTCGG GGCTGTACCA AAGGGCAGTC GGGCACCACA GGCCCGGGCC 17220
TCCACCTCAA CAGGCCTCCC GAGCCACTGG GAGCTGAATG CCAGGAGGCC GAAGCCCTCG CCCCATGAGG 17290
GCTGAGAAGG AGTGTGAGCA TTTGTGTTAC CCAGGGCCGA GGCTGCGCGA ATTACCGTGC ACACTTGATG 17360
TGAAATGAGG TCGTCGTCTA TCGTGGAAAC CCAGCAAGGG CTCACGGGAG AGTTTTCCAT TACAAGGTCG 17430
TACCATGAAA ATGGTTTTTA ACCCGAGTGC TTGCGCCTTC ATGCTCTGGC AGGGAGGGCA GAGCCACAGC 17500
TGCATGTTAC CGCCTTTGCA CCAGCTCCAG AGGCTTGGGA CCAGGCTGTC TCAGTTCCAG GGTGCGTCCG 17570
GCTCAGACCG CCCTCCTCTC TGCCTTCTCT CTCTGCCTCA AATCTTCCCT CGTTTGCATC TCCCTGACGC 17640
GTGCCTGGGC CCTCGTGCAA GCTGCTTGAC TCCTTTCCGG AAACCCTTGG GGTGTGCTGG ATACAGGTGC 17710
CACTGAGGAC TGGAGGTGTC TGACACTGTG GTTGACCCCA GGGTCCAGCT GGCGTGCTTG GGGCCTCCTT 17780
GGGCCATGAT GAGGTCAGAG GAGTTTTCCC AGGTGAAAAC TCCTGGGAAA CTCCCAGGGC CATGTGACCT 17850
GCCACCTGCT CCTCCCATAT TCAGCTCAGT CTTGTCCTCA TTTCCCCACC AGGGTCTCTA GCTCCGAGGA 17920
GCTCCCGTAG AGGGCCTGGG CTCAGGGCAG GGCGGCTGAG TTTCCCCACC CATGTGGGGA CCCTTGGGTA 17990
GTCGCTTGAT TGGGTAGCCC TGAGGAGGCC GAGATGCGAT GGGCCACGGG CCGTTTCCAA ACACAGAGTC 18060
AGGCACGTGG AAGGCCCAGG AATCCCCTTC CCTCGAGGCA GGAGTGGGAG AACGGAGAGC TGGGCCCCGA 18130
TTTCACGGCA GCCAGGCTGC AGTGGGCGAG GCTGTGGTGG TCCACGTGGC GCTGGGGGCG GGGTCTGATT 18200
CAAATCCGCT GGGGCTCGGC CTTCCTGGCC CGTGCTGGCC GCGCCTCCAC ACGGGCTTGG GGTGGACGCC 18270
CCGACCTCTA GCAGGTGGCT ATTTCTCCCT TTGGAAGAGA GCCCCTCACC CATGCTAGGT GTTTCCCTCC 18340
TGGGTCAGGA GCGTGGCCGT GTGGCAACCC CGGGACCTTA GGCTTATTTA TTTGTTTAAA AACATTCTGG 18410
GCCTGGCTTC CGTTGTTGCT AAATGGGGAA AAGACATCCC ACCTCAGCAG AGTTACTGAG AGGCTGAAAC 18480
CGGGGTGCTG GCTTGACTGG TGTGATCTCA GGTCATTCCA GAAGTGGCTC AGGAAGTCAG TGAGACCAGG 18550
TACATGGGGG GCTCAGGCAG TGGGTGAGAT GAGGTACACG GGGGGCTCAG GCAGTGGGTG AGGCCAGGTA 18620
CATGGGGGGC TCAGGCACTG GGTGAGATGA GGTACACGGG GGGCTCAGGC AGAGGGTCAG ACCAGGTACA 18690
CGGGGGCTCT GATCACACGC ACATATGAGC ACATGTGCAC ATGTGCTGTT TCATGGTAGC CAGGTCTGTG 18760
CACACCTGCC CCAAAGTCCC AGGAAGCTGA GAGGCCAAAG ATGGAGGCTG ACAGGGCTGG CGCGGTGGCT 18830
CACACCTGTA GTCCCAGCAC TTTGGGAGGC CGAGGCGAGA GGATCCCTTG AGCCCAGGAG TTTAAGACCA 18900
GCCTGAGCAA CATAGTAGAA CCCCATCTCT ATGAAAAATA AAAACAAAAA TTAGCTGAAC ATGGTGGTGT 18970
GCGCCTGTAG TTCCAATACT TGGGAGGCTG AAGTGGGAGG ATCACTTGAG CCCAGGAGGT GGAAGCTGCA 19040
GTGAGCTGAG ATTGCACCAC TGTACTGCAG CCTGGGTGAC AGAGTGAGAG CCCATCTCAA CAACAACAAA 19110
GAAGACTGAC AAATGCAGTT TCTTGGAAAG AAACATTTAG TAGGAACTTA ACCTACACAC AGAAGCCAAG 19180
TCGGTGTCTC GGTGTCAGTG AGATGAGATG ATGGGTCCTC ACACCATCAC CCCAGACCCA GGGTTTATGC 19250
ACCACAGGGG CGGGTGGCTC AGAAGGGATG CGCAGGACGT TGATATACGA TGACATCAAG GTTGTCTGAC 19320
GAAGGGCAGG ATTCATGATA AGTACCTGCT GGTACACAAG GAACAATGGA TAAACTGGAA ACCTTAGAGG 19390
CCTTCCCGGA ACAGGGGCTA ATCAGAAGCC AGCATGGGGG GCTGGCATCC AGGATGGAGC TGCTTCAGCC 19460
TCCACATGCG TGTTCATACA GATGGTGCAC AGAAACGCAG TGTACCTGTG CACACACAGA CACGCAGCTA 19530
CTCGCACACA CAAGCACACA CACAGACATG CATGCATGCA TCCGTGTGTG TGCACCTGTG CCCATGAGGA 19600
AACCCATGCA TGTGCATTCA TGCACGCACA CAGGCACCGG TGGGCCCATG CCCACACCCA CGAGCACCGT 19670
CTGATTAGGA GGCCTTTCCT CTGACGCTGT CCGCCATCCT CTCAGGTTTC ACGCATGTGT GCTGCAGCTC 19740
CCATTTCATC AGCAAGTTTG GAAGAACCCC ACATTTTTCC TGCGCGTCAT CTCTGACACG GCCTCCCTCT 19810
GCTACTCCAT CCTGAAAGCC AAGAACGCAG GTATGTGCAG GTGCCTGGCC TCAGTGGCAG CAGTGCCTGC 19880
CTGCTGGTGT TAGTGTGTCA GGAGACTGAG TGAATCTGGG CTTAGGAAGT TCTTACCCCT TTTCGCATCA 19950
GGAAGTGGTT TAACCCAACC ACTGTCAGGC TCGTCTGCCC GCCCTCTCGT GGGGTGAGCA GAGCACCTGA 20020
TGGAAGGGAC AGGAGCTGTC TGGGAGCTGC CATCCTTCCC ACCTTGCTCT GCCTGGGGAA GCGCTGGGGG 20090
GCCTGGTCTC TCCTGTTTGC CCCATGGTGG GATTTGGGGG GCCTGGCCTC TCCTGTTTGC CCTGTGGTGG 20160
GATTGGGCTG TCTCCCGTCC ATGGCACTTA GGGCCCTTGT GCAAACCCAG GCCAAGGGCT TAGGAGGAGG 20230
CCAGGCCCAG GCTACCCCAC CCCTCTCAGG AGCAGAGGCC GCGTATCACC ACGACAGAGC CCCGCGCCGT 20300
CCTCTGCTTC CCAGTCACCG TCCTCTGCCC CTGGACACTT TGTCCAGCAT CAGGGAGGTT TCTGATCCGT 20370
CTGAAATTCA AGCCATGTCG AACCTGCGGT CCTGAGCTTA ACAGCTTCTA CTTTCTGTTC TTTCTGTGTT 20440
GTGGAAATTT CACCTGGAGA AGCCGAAGAA AACATTTCTG TCGTGACTCC TGCGGTGCTT GGGTCGGGAC 20510
AGCCAGAGAT GGAGCCACCC CGCAGACCGT CGGGTGTGGG CAGCTTTCCG GTGTCTCCTG GGAGGGGAGC 20580
TGGGCTGGGC CTGTGACTCC TCAGCCTCTG TTTTCCCCCA GGGATGTCGC TGGGGGCCAA GGGCGCCGCC 20650
GGCCCTCTGC CCTCCGAGGC CGTGCAGTGG CTGTGCCACC AAGCATTCCT GCTCAAGCTG ACTCGACACC 20720
GTGTCACCTA CGTGCCACTC CTGGGGTCAC TCAGGACAGG CAAGTGTGGG TGGAGGCCAG TGCGGGCCCC 20790
ACCTGCCCAG GGGTCATCCT TGAACGCCCT GTGTGGGGCG AGCAGCCTCA GATGCTGCTG AAGTGCAGAC 20860
GCCCCCGGGC CTGACCCTGG GGGCCTGGAG CCACGCTGGC AGCCCTATGT GATTAAACGC TGGTGTCCCC 20930
AGGCCACGGA GCCTGGCAGG GTCCCCAACT TCTTGAACCC CTGCTTCCCA TCTCAGGGGC GATGGCTCCC 21000
CACGCTTGGG AGCCTTCTGA CCCCTGACCT GTGTCCTCTC ACAGCCTCTT CCCTGGCTGC TGCCCTGAGC 21070
TCCTGGGGTC CTGAGCAAGT TCTCTCCCCG CCCCGCCGCT CCAGCGTCAC TGGGCTGCCT GTCTGCTCGC 21140
CCCGGTGGAG GGGTGTCTGT CCCTTCACTG AGGTTCCCAC CAGCCAGGGC CACGAGGTGC AGGCCCTGCC 21210
```

```
TGCCCGGCCA CCCACACGTC CTAGGAGGGT TGGAGGATGC CACCTCTGGC CTCTTCTGGA ACGGAGTCTG 21280
ATTTTGGCCC CGCAGCCCAG ACGCAGCTGA GTCGGAAGCT CCCGGGGACG ACGCTGACTG CCCTGGAGGC 21350
CGCAGCCAAC CCGGCACTGC CCTCAGACTT CAAGACCATC CTGGACTGAT GGCCACCCGC CCACAGCCAG 21420
GCCGAGAGCA GACACCAGCA GCCCTGTCAC GCCGGGCTCT ACGTCCCAGG GAGGGAGGGG CGGCCCACAC 21490
CCAGGCCCGC ACCGCTGGGA GTCTGAGGCC TGAGTGAGTG TTTGGCCGAG GCCTGCATGT CCGGCTGAAG 21560
GCTGAGTGTC CGGCTGAGGC CTGAGCGAGT GTCCAGCCAA GGGCTGAGTG TCCAGCACAC CTGCCGTCTT 21630
CACTTCCCCA CAGGCTGGCG CTCGGCTCCA CCCCAGGGCC AGCTTTTCCT CACCAGGAGC CCGGCTTCCA 21700
CTCCCCACAT AGGAATAGTC CATCCCCAGA TTCGCCATTG TTCACCCCTC GCCCTGCCCT CCTTTGCCTT 21770
CCACCCCCAC CATCCAGGTG GAGACCCTGA GAAGGACCCT GGGAGCTCTG GGAATTTGGA GTGACCAAAG 21840
GTGTGCCCTG TACACAGGCG AGGACCCTGC ACCTGGATGG GGGTCCCTGT GGGTCAAATT GGGGGGAGGT 21910
GCTGTGGGAG TAAAATACTG AATATATGAG TTTTTCAGTT TTGAAAAAAA TCTCATGTTT GAATCCTAAT 21980
GTGCACTGCA TAGACACCAC TGTATGCAAT TACAGAAGCC TGTGAGTGAA CGGGGTGGTG GTCAGTGCGG 22050
GCCCATGGCC TGGCTGTGCA TTTACGGAAG TCTATGAGTG AATGGGGTTG TGGTCAGTGC GGGCCCATGG 22120
CCTGGCTGGG CCTGGGAGGT TTCTGATGCT GTGAGGCAGG AGGGGAAGGA GGGTAGGGGA TAGACAGTGG 22190
GAGCCCCCAC CCTGGAAGAC ATAACAGTAA GTCCAGGCCC GAAGGGCAGC AGGGATGCTG GGGGCCCAGC 22260
TTGGGCGGCG GGGATGATGG AGGGCCTGGC CAGGGTGGCA GGGATGATGG GGGCCCCAGC TGGGGTGGCA 22330
GGGGTGATGG GGGGGGCTGG TCTGGGTGGC GGGGAAGATG GGGAAGCCTG GCTGGGCCCC CTCCTCCCCT 22400
GCCTCCCACC TGCAGCCGTG GATCCGGATG TGCTTCCCTG GTGCACATCC TCTGGGCCAT CAGCTTTCAT 22470
GGAGGTGGGG GGCAGGGGCA TGACACCATC CTGTATAAAA TCCAGGATTC CTCCTCCTGA ACGCCCCAAC 22540
TCAGGTTGAA AGTCACATTC CGCCTCTGGC CATTCTCTTA AGAGTAGACC AGGATTCTGA TCTCTGAAGG 22610
GTGGGTAGGG TGGGGCAGTG GAGGGTGTGG ACACAGGAGG CTTCAGGGTG GGGCTGGTGA TGCTCTCTCA 22680
TCCTCTTATC ATCTCCCAGT CTCATCTCTC ATCCTCTTAT CATCTCCCAG TCTCATCTGT CTTCCTCTTA 22750
TCTCCCAGTC TCATCTGTCA TCCTCTTACC ATCTCCCAGT CTCATCTCTT ATCCTCTTAT CTCCTAGTCT 22820
CATCCAGACT TACCTCCCAG GGCGGGTGCC AGGCTCGCAG TGGAGCTGGA CATACGTCCT TCCTCAGGCA 22890
GAAGGAACTG GAAGGATTGC AGAGAACAGG AGGGGCGGCT CAGAGGGACG CAGTCTTGGG GTGAAGAAAC 22960
AGCCCCTCCT CAGAAGTTGG CTTGGGCCAC ACGAAACCGA GGGCCCTGCG TGAGTGGCTC CAGAGCCTTC 23030
CAGCAGGTCC CTGGTGGGGC CTTATGGTAT GGCCGGGTCC TACTGAGTGC ACCTTGGACA GGGCTTCTGG 23100
TTTGAGTGCA GCCCGGACGT GCCTGGTGTC GGGGTGGGGG CTTATGGCCA CTGGATATGG CGTCATTTAT 23170
TGCTGCTGCT TCAGAGAATG TCTGAGTGAC CGAGCCTAAT GTGTATGGTG GGCCCAAGTC CACAGACTGT 23240
GTCGTAAATG CACTCTGGTG CCTGGAGCCC CCGTATAGGA GCTGTGAGGA AGGAGGGGCT CTTGGCAGCC 23310
GGCCTGGGGG CGCCTTTGCC CTGCAAACTG GAAGGGAGCG GCCCCGGGCG CCGTGGGCGG ACGACCTCAA 23380
GTGAGAGGTT GGACAGAACA GGGCGGGGAC TTCCCAGGAG CAGAGGCCGC TGCTCAGGCA CACCTGGGTT 23450
TGAATCACAG ACCAACaGGT CAGGCCATTG TTCAGCTATC CATCTTCTAC AAAGCTCCAG ATTCCTGTTT 23520
CTCCGGGTGT TTTTTGTTGA AATTTTACTC AGGATTACTT ATATTTTTTG CTAAAGTATT AGACCCTTAA 23590
AAAAGGTATT TGCTTTGATA TGGCTTAACT CACTAAGCAC CTACTTTATT TGTCTGTTTT TATTTATTAT 23660
TATTATTATT ATTAGAGATG GTGTCTACTC TGTCACCCAG GTTGTTAGTG CAGTGGCACA GTCATGGCTC 23730
GCTGTAGCCG CAAACCCCCA GGCTCAAGTG ATCCTCCGGC CTCAGCTTCC CAGAGTGCTG GGATTACAGG 23800
TGTGAGCCAC TGCCCTTGCC TGGCACTTTT AAAAACCACT ATGTAAGGTC AGGTCCAGTG GCTTCCACAC 23870
CTGTCATCCC AGTAGTTTGG GAAGCCGAGG CAGAAGGATT GTCTGAGGCC AGGAGTTTGA GACCAGCATG 23940
GGTAACATAG GGAGACCCCA TCTCTACAAA AAATGCAAAA AGTTATCCGG GCGTGGGGTC CAGCATCTGT 24010
AGTCCCAGCT GCTCGGGAGG CTGAGTGGGA GGATCGCTTG AGCCCGGGAG GTCATGGCTG CAGTGAGCTG 24080
TGATTGTACC ATCGCACTCC AGCCTGGGCA ACAGAGTGAG ACCCTGTCTC AAAAAAAAAA AAAAAAAAAG 24150
AAGGAGAAGG AGAAGAGAAG AAGAAGGAAG AAGGAAAGAG AAGAAGAAGG AAGAAGGAAG AAAGAAGGAG 24220
AAGGAGGCCT GCTAGGTGCT AGGTAGACTG TCAAATCTCA GAGCAAAATG AAAATAACAA AGTTTTAAAG 24290
GGAAAGAAAA ACCCCAGCTC TTTGGACTTC CTTAGGCCTG AACTTCATCT CAAGCAGCTT CCTTCCACAG 24360
ACAAGCGTGT ATGGAGCGAG TGAGTTCAAA GCAGAAAGGG AGGAGAAGCA GGCAAGGGTG GAGGCTGTGG 24430
GTGACACCAG CCAGGACCCC TGAAAGGGAG TGGTTGTTTT CCTGCCTCAG CCCCACGCTC CTGCCGGTCC 24500
TGCACCTGCT GTAACCGTCG ATGTTGGTGC CAGGTGCCCA CCTGGGAAGG ATGCTGTGCA GGGGGCTTGC 24570
CAAACTTTGG TGGGTTTCAG AAGCCCCAGG CACTTGTGGC AGGCACAATT ACAGCCCCTC CCCAAAGATG 24640
CCCACGTCCT TCTCCTGGAA CCTGTGAATG TGTCACCCGC AAGGCAGAGG CTGGTGAAGG CTGCAGGTGG 24710
AATCACGGCT GCCAGTCAGC CGATCTTAAG GTCATCCTGG ATTATCTGGT GGGCCTGATA TGGCCACAAG 24780
GGTCCCTAGA AGTGAGAGAG GGAGGCAGGG GAGAGTCAGA GAGGGGACGT GAGAAGGACC ACTGGCCACT 24850
GCTGGCTTTG AGATGGAGGA GGGGGTCCCC AGCCAAGGAA TGGGGGCAGC CGCTCCATGC TGGAAAAGCA 24920
AGCAATCCTC CCCGGTCCTG AGGGCACACG GCCCTGCCCA CGCCTCGATT TCAGGCCAGT GGGACCTGTT 24990
TCAGCTTTCC GGCCTCCAGA GCTGTAAGAT GATGCGTTTG TGTTCAGCCA CTAAGCTGCA GTGATTCGTC 25060
ACAGCAGCAA ATGGAATAGC AGTACAGGGA AATGAATACA GGGACAGTTC TCAGAGTGAC TCTCAGCCCA 25130
CCCCTGGG                                                                    25138
```

Beispiel 5

Der Vergleich der oben beschriebenen genomischen hTC-Sequenz mit der Sequenz der hTC-cDNA (Fig. 6; entsprechend SEQ ID NO 2) ermöglichte die Aufklärung der Exon-Intron-Struktur des hTC-Gens. Die genomische Organisation des hTC-Gens ist in Fig. 7 schematisch dargestellt. Die kodierende Region des hTC-Gens setzt sich aus

16 Exons zusammen, die in ihrer Größe zwischen 62 bp und 1354 bp variieren (s. Tabelle 1). Exon 1 enthält das Translationsstartcodon ATG. Das Translationsstopcodon TGA sowie der 3´ untranslatierte Bereich liegen auf Exon 16 (Fig. 8). Ein mögliches Polyadenylierungssignal (AATAAA) wurde weder in Exon 16 noch in den 3195 bp der folgenden 3'-flankierenden Region gefunden. Basierend auf der Konsensussequenz

	5'-Exon			Intron							3'-Exon		
Prä-mRNA	A/C	A	G	G	T	A/G	A	...	N	C	A	G	G
Häufigk.(%)	70	60	80	100	100	95	70		80	100	100	100	60

wurden die Exon-Intron-Übergänge bestimmt und in Tabelle 1 aufgeführt. Mit Ausnahme der 5'-Splice-Stelle zwischen Exon 15 und Intron 15 stimmen alle Exon-Intron-Übergänge mit der publizierten (Shapiro und Senapathy, 1987) Splice-Konsensussequenz überein. Die Größe der Introns liegt zwischen 104 bp und 8616 bp. Da Intron 6 nur zum Teil isoliert wurde, kann die exakte Länge des hTC-Gens nicht bestimmt werden. Basierend auf der von Intron 6 erhaltenen Teilsequenz von ~4660 bp beträgt die minimale Größe des hTERT Gens 37 kb.

Die Introns 1-5 sowie der 5'-Bereich des Introns 6 sind in Contig 1 enthalten:

Intron 1: bp 11493-11596 (SEQ ID NO 4);

Intron 2: bp 12951-21566 (SEQ ID NO 5);

Intron 3: bp 21763-23851 (SEQ ID NO 6);

Intron 4: bp 24033-24719 (SEQ ID NO 7);

Intron 5: bp 24900-25393 (SEQ ID NO 8);

5'-Bereich von Intron 6: bp 25550-26414 (SEQ ID NO 9).

Der 3'-Bereich des Introns 6 sowie die Introns 7-15 sind in Contig 2 an folgenden Positionen lokalisiert:

3'-Bereich von Intron 6: bp 1-3782 (SEQ ID NO 10);

Intron 7: bp 3879-4858 (SEQ ID NO 11);

Intron 8: bp 4945-7429 (SEQ ID NO 12);

Intron 9: bp 7544-9527 (SEQ ID NO 13);

Intron 10: bp 9600-11470 (SEQ ID NO 14);

Intron 11: bp 11660-15460 (SEQ ID NO 15;

Intron 12: bp 15588-16467 (SEQ ID NO 16);

Intron 13: bp 16530-19715 (SEQ ID NO 17);

Intron 14: 19841-20621 (SEQ ID NO 18);

Intron 15: 20760-21295 (SEQ ID NO 19).

Der 3'-nichttranskribierte Bereich befindet sich ebenfalls im Contig 2 an Position 21960-25138 (SEQ ID NO 20).

Die genannten Introns haben im einzelnen folgende Sequenzen:

Intron 1 (SEQ ID NO 4)

GTGGGCCTCCCCGGGGTCGGCGTCCGGCTGGGGTTGAGGGCGGCCGGGGGGAACCAGCGACATGCGGAGAGCAGCGCAGG
CGACTCAGGGCGCTTCCCCCGCAG

Intron 2 (SEQ ID NO 5)

GTGAGGAGGTGGTGGCCGTCGAGGGCCCAGGCCCCAGAGCTGAATGCAGTAGGGGCTCAGAAAAGGGGGCAGGCAGAGCC
CTGGTCCTCCTGTCTCCATCGTCACGTGGGCACACGTGGCTTTTCGCTCAGGACGTCGAGTGGACACGGTGATCTCTGCC
TCTGCTCTCCCTCCTGTCCAGTTTGCATAAACTTACGAGGTTCACCTTCACGTTTTGATGGACACGCGGTTTCCAGGCGC
CGAGGCCAGAGCAGTGAACAGAGGAGGCTGGGCGCGGCAGTGGAGCCGGGTTGCCGGCAATGGGGAGAAGTGTCTGGAAG
CACAGACGCTCTGGCGAGGGTGCCTGCAGGTTACCTATAATCCTCTTCGCAATTTCAAGGGTGGGAATGAGAGGTGGGGA
CGAGAACCCCCTCTTCCTGGGGGTGGGAGGTAAGGGTTTTGCAGGTGCACGTGGTCAGCCAATATGCAGGTTTGTGTTTA
AGATTTAATTGTGTGTTGACGGCCAGGTGCGGTGGCTCACGCCGGTAATCCCAGCACTTTGGGAAGCTGAGGCAGGTGGA
TCACCTGAGGTCAGGAGTTTGAGACCAGCCTGACCAACATGGTGAAACCCTATCTGTACTAAAAATACAAAAATTAGCTG
GGCATGGTGGTGTGTGCCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGAGGC
TGCAGTGAGCTGAGATTGTGCCATTGTACTCCAGCCTGGGCGACAAGAGTGAAACTCTGTCTTTAAAAAAAAAAAAGTGTT
CGTTGATTGTGCCAGGACAGGGTAGAGGGAGGGAGATAAGACTGTTCTCCAGCACAGATCCTGGTCCCATCTTTAGGTAT
GAAGAGGGCCACATGGGAGCAGAGGACAGCAGATGGCTCCACCTGCTGAGGAAGGGACAGTGTTTGTGGGTGTTCAGGGG
ATGGTGCTGCTGGGCCCTGCCGTGTCCCCACCCTGTTTTTCTGGATTTGATGTTGAGGAACCTCCGCTCCAGCCCCCTTT
TGGCTCCCAGTGCTCCCAGGCCCTACCGTGGCAGCTAGAAGAAGTCCCGATTTCACCCCCTCCCCACAAACTCCCAAGAC
ATGTAAGACTTCCGGCCATGCAGACAAGGAGGGTGACCTTCTTGGGGCTCTTTTTTTTCTTTTTTTCTTTTTATGGTGGC
AAAAGTCATATAACATGAGATTGGCACTCCTAACACCGTTTTCTGTGTACAGTGCAGAATTGCTAACTCGGCGGTGTTTA
CAGCAGGTTGCTTGAAATGCTGCGTCTTGCGTGACTGGAAGTCCCTACCCATCGAACGGCAGCTGCCTCACACCTGCTGC
GGCTCAGGTGGACCACGCCGAGTCAGATAAGCGTCATGCAACCCAGTTTTGCTTTTTGTGCTCCAGCTTCCTTCGTTGAG
GAGAGTTTGAGTTCTCTGATCAGGACTCTGCCTGTCATTGCTGTTCTCTGACTTCAGATGAGGTCACAATCTGCCCCTGG
CTTATGCAGGGAGTGAGGCGTGGTCCCCGGGTGTCCCTGTCACGTGCAGGGTGAGTGAGGCGTTGCCCCCAGGTGTCCCT
GTCACGTGTAGGGTGAGTGAGGCGCGGCCCCCGGGTGTCCCTGTCCCGTGCAGCGTGATTGAGGTGTGGCCCCCGGGTGT
CCCTGTCACGTGTAGGGTGAGTGAGGCGCCATCCCCGGGTGTCCCTGTCACGTGTAGGGTGAGTGAGGCGTGGTCCCCGG
GTGTCCCTGTCCCGTGCAGGGTGAGTGAGGCACTGTCCCCGGGTGTCCCTGTCACGTGCAGGGTGAGTGAGGCGCGGTCC
CCGGGTGTCCCTCTCAGGTGTAGGGTGAGTGAGGCGCGGCCCCAGGGTGTCCCTGTCACGTGTAGGGTGAGTGAGGCACC
GTCCCTGGGTGTCCCTCCCAGGTATAGGGTGAGTGAGGCACTGTCCCCGGGTGTCCCTGTCACGTGCAGGGTGAGTGAGG
CGCGGCCCCCGGGTGTCCCTCTCAGGTGCAGGGTGAGTGAGGCGCTGTCCCTGGGTGTCCCTGTCTCGTGTAGGGTGAGT
GAGGCTCTGTCCCCAGGTGTCCTTGGCGTTTGCTCACTTGAGCTTGCTCCTGAATGTTTGCTCTTTCTATAGCCACAGCT
GCGCCGGTTGCCCATTGCCTGGGTAGATGGTGCAGGCGCAGTGCTGGTCCCCAAGCCTATCTTTTCTGATGCTCGGCTCT
TCTTGGTCACCTCTCCGTTCCATTTTGCTACGGGGACACGGGACTGCAGGCTCTCGCCTCCCGCGTGCCAGGCACTGCAG
CCACAGCTTCAGGTCCGCTTGCCTCTGTTGGGCCTGGCTTGCTCACCACGTGCCCGCCACATGCATGCTGCCAATACTCC
TCTCCCAGCTTGTCTCATGCCGAGGCTGGACTCTGGGCTGCCTGTGTCTGCTGCCACGTGTTGCTGGAGACATCCCAGAA
AGGGTTCTCTGTGCCCTGAAGGAAAGCAAGTCACCCCAGCCCCCTCACTTGTCCTGTTTTCTCCCAAGCTGCCCCTCTGC
TTGGCCCCCTTGGGTGGGTGGCAACGCTTGTCACCTTATTCTGGGCACCTGCCGCTCATTGCTTAGGCTGGGCTCTGCCT
CCAGTCGCCCCCTCACATGGATTGACGTCCAGCCACAGGTTGGAGTGTCTCTGTCTGTCTCCTGCTCTGAGACCCACGTG

GAGGGCCGGTGTCTCCGCCAGCCTTCGTCAGACTTCCCTCTTGGGTCTTAGTTTTGAATTTCACTGATTTACCTCTGACG
TTTCTATCTCTCCATTGTATGCTTTTTCTTGGTTTATTCTTTCATTCCTTTTCTAGCTTCTTAGTTTAGTCATGCCTTTC
CCTCTAAGTGCTGCCTTACCTGCACCCTGTGTTTTGATGTGAAGTAATCTCAACATCAGCCACTTTCAAGTGTTCTTAAA
ATACTTCAAAGTGTTAATACTTCTTTTAAGTATTCTTATTCTGTGATTTTTTTCTTTGTGCACGCTGTGTTTTGACGTGA
AATCATTTTGATATCAGTGACTTTTAAGTATTCTTTAGCTTATTCTGTGATTTCTTTGAGCAGTGAGTTATTTGAACACT
GTTTATGTTCAAGATATGTAGAGTATCAAGATACGTAGAGTATTTTAAGTTATCATTTTATTATTGATTTCTAACTCAGT
TGTGTAGTGGTCTGTATAATACCAATTATTTGAAGTTTGCGGAGCCTTGCTTTGTGATCTAGTGTGTGCATGGTTTCCAG
AACTGTCCATTGTAAATTTGACATCCTGTCAATAGTGGGCATGCATGTTCACTATATCCAGCTTATTAAGGTCCAGTGCA
AAGCTTCTGTCTCCTTCTAGATGCATGAAATTCCAAGAAGGAGGCCATAGTCCCTCACCTGGGGGATGGGTCTGTTCATT
TCTTCTCGTTTGGTAGCATTTATGTGAGGCATTGTTAGGTGCATGCACGTGGTAGAATTTTTATCTTCCTGATGAGTGAA
TCTTTTGGAGACTTCTATGTCTCTAGTAATCTAGTAATTCTTTTTTTAAATTGCTCTTAGTACTGCCACACTGGGCTTCT
TTTGATTAGTATTTTCCTGCTGTGTCTGTTTTCTGCCTTTAATTTATATATATATATATATTTTTTTTTTTTTTGAGACA
GAGTCTTGGTCTGTCGCCCAGGGTGAGTGCAGTGGTGTGATCACAGGTCAGTGTAACTTTTACCTTCTGGCCTGAGCCGT
CCTCTCACCTCAGCCTCCTGAGTAGCTGGAACTGCAGACACGCACCGCTACACCTGGCTAATTTTTAAATTTTTTCTGGA
GACAGGGTCTTGCTGTGTTGCCCAGGCTGGTCTCAAACTCTTGGACTCAAGGGATCCATCTACCTCGGCTTCCCAAAGTG
CTGAATTACAGGCATGAGCCACCATGTCTGGCCTAATTTTCAACACTTTTATATTCTTATAGTGTGGGTATGTCCTGTTA
ACAGCATGTAGGTGAATTTCCAATCCAGTCTGACAGTCGTTGTTTAACTGGATAACCTGATTTATTTTCATTTTTTTGTC
ACTAGAGACCCGCCTGGTGCACTCTGATTCTCCACTTGCCTGTTGCATGTCCTCGTTCCCTTGTTTCTCACCACCTCTTG
GGTTGCCATGTGCGTTTCCTGCCGAGTGTGTGTTGATCCTCTCGTTGCCTCCTGGTCACTGGGCATTTGCTTTTATTTCT
CTTTGCTTAGTGTTACCCCCTGATCTTTTTATTGTCGTTGTTTGCTTTTGTTTATTGAGACAGTCTCACTCTGTCACCCA
GGCTGGAGTGTAATGGCACAATCTCGGCTCACTGCAACCTCTGCCTCCTCGGTTCAAGCAGTTCTCATTCCTCAACCTCA
TGAGTAGCTGGGATTACAGGCGCCCACCACCACGCCTGGCTAATTTTTGTATTTTTAGTAGAGATAGGCTTTCACCATGT
TGGCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCCGCCTTGGCCTCCCACAGTGCTGGGATTACAGGTGCAA
GCCACCGTGCCCGGCATACCTTGATCTTTTAAAATGAAGTCTGAAACATTGCTACCCTTGTCCTGAGCAATAAGACCCTT
AGTGTATTTTAGCTCTGGCCACCCCCCAGCCTGTGTGCTGTTTTCCCTGCTGACTTAGTTCTATCTCAGGCATCTTGACA
CCCCCACAAGCTAAGCATTATTAATATTGTTTTCCGTGTTGAGTGTTTCTGTAGCTTTGCCCCCGCCCTGCTTTTCCTCC
TTTGTTCCCCGTCTGTCTTCTGTCTCAGGCCCGCCGTCTGGGGTCCCCTTCCTTGTCCTTTGCGTGGTTCTTCTGTCTTG
TTATTGCTGGTAAACCCCAGCTTTACCTGTGCTGGCCTCCATGGCATCTAGCGACGTCCGGGGACCTCTGCTTATGATGC
ACAGATGAAGATGTGGAGACTCACGAGGAGGGCGGTCATCTTGGCCCGTGAGTGTCTGGAGCACCACGTGGCCAGCGTTC
CTTAGCCAGTGAGTGACAGCAACGTCCGCTCGGCCTGGGTTCAGCCTGGAAAACCCCAGGCATGTCGGGGTCTGGTGGCT
CCGCGGTGTCGAGTTTGAAATCGCGCAAACCTGCGGTGTGGCGCCAGCTCTGACGGTGCTGCCTGGCGGGGGAGTGTCTG
CTTCCTCCCTTCTGCTTGGGAACCAGGACAAAGGATGAGGCTCCGAGCCGTTGTCGCCCAACAGGAGCATGACGTGAGCC
ATGTGGATAATTTTAAAATTTCTAGGCTGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGCGGG
TGGATCACGAGGTCAGGAGGTCGAGACCATCCTGGCCAACATGATGAAACCCCATCTGTACTAAAAACACAAAAATTAGC
TGGGCGTGGTGGCGGGTGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGTTGGAA
GTTGCAGTGAGCCGACATTGCACCACTGCACTCCAGCCTGGCAACACAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAA
AAAAAAAAAAAATTCTAGTAGCCACATTAAAAAAGTAAAAAAGAAAAGGTGAAATTAATGTAATAATAGATTTTACTGAA
GCCCAGCATGTCCACACCTCATCATTTTAGGGTGTTATTGGTGGGAGCATCACTCACAGGACATTTGACATTTTTTGAGC
TTTGTCTGCGGGATCCCGTGTGTAGGTCCCGTGCGTGGCCATCTCGGCCTGGACCTGCTGGGCTTCCCATGGCCATGGCT
GTTGTACCAGATGGTGCAGGTCCGGGATGAGGTCGCCAGGCCCTCAGTGAGCTGGATGTGCAGTGTCCGGATGGTGCACG
TCTGGGATGAGGTCGCCAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTCAGGGGTGAGGTCTCCAG

GCCCTCGGTGAGCTGGAGGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTCGCCAGGCCCTGCTGTGAGCTGGATG
TGTGGTGTCTGGATGGTGCAGGTCAGGGGTGAGGTCTCCAGGCCCTCGGTAAGCTGGAGGTATGGAGTCCGGATGATGCA
GGTCCGGGGGTGAGGTCGCCAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTCTGGGGTGAGGTCACC
AGGCCCTGCGGTGAGCTGGGTGTGCGGTGTCTGGATGGTGCAGGTCTGGAGTGAGGTCGCCAGACGGTGCCAGACCATGC
GGTGAGCTGGATATGCGGTGTCCGGATGGTGCAGGTCTGGGGTGAGGTTGCCAGGCCCTGCTGTGAGTTGGATGTGGGGT
GTCCGGATGCTGCAGGTCCGGTGTGAGGTCACCAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTCT
GGGGTGAAGGTCGCCAGGCCCCTGCTTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTCTGGAGTGAGGTCGCCAG
GCCCTCGGTGAGCTGGATGTGCAGTGTCCAGATGGTGCAGGTCCGGGGTGAGGTCGCCAGACCCTGCGGTGAGCTGGATG
TGCGGTGTCTGGATGGTGCAGGTCTGGAGTGAGGTCGCCAGGCCCTCGGTGAGCTGGATGTATGGAGTCCGGATGGTGCC
GGTCCGGGGTGAGGTCGCCAGACCCTGCTGTGAGCTGGATGTGCGGTGTCTGGATGGTACAGGTCTGGAGTGAGGTCGCC
AGACCCTGCTGTGAGCTGGATATGCGGTGTCCGGATGGTGCAGGTCAGGGGTGAGGTCTCCAGGCCCTCGGTGAGCTGGA
GGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTCGCCAGGCCCTGCTGTGAACTGGATGTGCGGCGTCTGGATGGT
GCAGGTCTGGGGTGTGGTCGCCAGGCCCTCGGTGAGCTGGAGGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTCG
CCAGGCCCTGCTGTGAGCTGGATGTGCGGCGTCTGGATGGTGCAGGTCTGGGGTGTGGTCGCCAGGCCCTCGGTGAGCTG
GAGGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTTGCCAGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATG
GTGCAGTCCGGGGTGAGGTCGCCAGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTGCAGGTCTGGGGTGAGGT
CACCAGGCCCTGCGGTGAGCTGGTTGTGCGGTGTCCGGTTGCTGCAGGTCCGGGGTGAGTTCGCCAGGCCCTCGGTGAGC
TGGATGTGCGGTGTCCCCGTGTCCGGATGGTGCAGGTCCAGGGTGAGGTCGCTAGGCCCTTGGTGGGCTGGATGTGCCGT
GTCCGGATGGTGCAGGTCTGGGGTGAGGTCGCCAGGCCTTTGGTGAGCTGGATGTGCGGTGTCTGCATGGTGCAGGTCTG
GGGTGAGGTCGCCAGGCCCTTGGTGGGCTGGATGTGTGGTGTCCGGATGGTGCAGGTCCGGCGTGAGGTCGCCAGGCCCT
GCTGTGAGCTGGATGTGCGGTGTCTGGATGGTGCAGGTCCGGGGTGAGGTAGCCAAGGCCTTCGGTGAGCTGGATGTGGG
GTGTCCGGATGGTGCAGGTCCGGGGTGAGGTCGCCAGGCCCTGCGGTTAGCTGGATATGCGGTGTCCGGATGGTGCAGGT
CCGGGGTGAGGTCACCAGGCCCTGCGGTTAGCTGGATGTGCGGTGTCTGGATGGTGCAGGTCCGGGGTGAGGTCGCCAGG
CCCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTGCAGGTCCGGGGTGAGGTCGCCAGGCCCTGCAGTGAGCTGGATG
TGCTGTATCCGGATGGTGCAGGTCTGGCGTGAGGTCGCCAGGCCCTGCGGTTAGCTGGATATGCGGTGTCGGATGGTGCA
GGTCCGGGGTGAGGTCACCAGGCCCTGCGGTTAGCTGGATGTGCGGTGTCCGGATGGTGCAGGTCTGGGGTGAGGTCGCC
AGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTGCAGGTCCGGGGTGAGGTCGCCAGGCCCTGCGGTGAGCTGG
ATGTGCTGTATCCGGATGGTGCAGGTCTGGCGTGAGGTCGCCAGGCCCTGCGGTGAGCTGGATGTGCAGTGTACGGATGG
TGCAGGTCCGGGGTGAGGTCGCCAGGCCCTGCGGTGGGCTGTATGTGTGTTGTCTGGATGGTGCAGGTCCGGGGTGAGTT
CGCCAGGCCCTGCGGTGAGCTGGATGTGTGGTGTCTGGATGCTGCAGGTCCGGGGTGAGTTCGCCAGGCCCTCGGTGAGC
TGGATATGCGGTGTCCCCGTGTCCGAATGGTGCAGGTCCAGGGTGAGGTCGCCAGGCCCTTGGTGGGCTGGATGTGCCGT
GTCCGGATGGTGCAGGTCTGGGGTGAGGTCGCCAGGCCCTTGGTGAGCTGGATGTGCGGTGTCCGGATGGTGCAGGTCCG
GGGTGAGGTCACCAGGCCCTCGGTGATCTGGATGTGGCATGTCCTTCTCGTTTAAG

Intron 3 (SEQ ID NO 6)

GTACTGTATCCCCACGCCAGGCCTCTGCTTCTCGAAGTCCTGGAACACCAGCCCGGCCTCAGCATGCGCCTGTCTCCACT
TGCCTGTGCTTCCCTGGCTGTGCAGCTCTGGGCTGGGAGCCAGGGGCCCCGTCACAGGCCTGGTCCAAGTGGATTCTGTG
CAAGGCTCTGACTGCCTGGAGCTCACGTTCTCTTACTTGTAAAATCAGGAGTTTGTGCCAAGTGGTCTCTAGGGTTTGTA
AAGCAGAAGGGATTTAAATTAGATGGAAACACTACCACTAGCCTCCTTGCCTTTCCCTGGGATGTGGGTCTGATTCTCTC
TCTCTTTTTTTTTTCTTTTTTGAGATGGAGTCTCACTCTGTTGCCCAGGCTGGAGTGCAGTGGCATAATCTTGGCTCACT

GCAACCTCCACCTCCTGGGTTTAAGCGATTCACCAGCCTCAGCCTCCTAAGTAGCTGGGATTACAGGCACCTGCCACCAC
GCCTGGCTAATTTTTGTACTTTTAGGAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCATGACCTCAGG
TGATCCACCCACCTTGGCCTCCCAAAGTGCTGGGTTTACAGGCTAAGCCACCGTGCCCAGCCCCCGATTCTCTTTTAATT
CATGCTGTTCTGTATGAATCTTCAATCTATTGGATTTAGGTCATGAGAGGATAAAATCCCACCCACTTGGCGACTCACTG
CAGGGAGCACCTGTGCAGGGAGCACCTGGGGATAGGAGAGTTCCACCATGAGCTAACTTCTAGGTGGCTGCATTTGAATG
GCTGTGAGATTTTGTCTGCAATGTTCGGCTGATGAGAGTGTGAGATTGTGACAGATTCAAGCTGGATTTGCATCAGTGAG
GGACGGGAGCGCTGGTCTGGGAGATGCCAGCCTGGCTGAGCCCAGGCCATGGTATTAGCTTCTCCGTGTCCCGCCCAGGC
TGACTGTGGAGGGCTTTAGTCAGAAGATCAGGGCTTCCCCAGCTCCCCTGCACACTCGAGTCCCTGGGGGGCCTTGTGAC
ACCCCATGCCCCAAATCAGGATGTCTGCAGAGGGAGCTGGCAGCAGACCTCGTCAGAGGTAACACAGCCTCTGGGCTGGG
GACCCCGACGTGGTGCTGGGGCCATTTCCTTGCATCTGGGGGAGGGTCAGGGCTTTCCCTGTGGGAACAAGTTAATACAC
AATGCACCTTACTTAGACTTTACACGTATTTAATGGTGTGCGACCCAACATGGTCATTTGACCAGTATTTTGGAAAGAAT
TTAATTGGGGTGACCGGAAGGAGCAGACAGACGTGGTGGTCCCCAAGATGCTCCTTGTCACTACTGGGACTGTTGTTCTG
CCTGGGGGGCCTTGGAGGCCCCTCCTCCCTGGACAGGGTACCGTGCCTTTTCTACTCTGCTGGGCCTGCGGCCTGCGGTC
AGGGCACCAGCTCCGGAGCACCCGCGGCCCCAGTGTCCACGGAGTGCCAGGCTGTCAGCCACAGATGCCCAGGTCCAGGT
GTGGCCGCTCCAGCCCCCGTGCCCCCATGGGTGGTTTTGGGGGAAAAGGCCAAGGGCAGAGGTGTCAGGAGACTGGTGGG
CTCATGAGAGCTGATTCTGCTCCTTGGCTGAGCTGCCCTGAGCAGCCTCTCCCGCCCTCTCCATCTGAAGGGATGTGGCT
CTTTCTACCTGGGGGTCCTGCCTGGGGCCAGCCTTGGGCTACCCCAGTGGCTGTACCAGAGGGACAGGCATCCTGTGTGG
AGGGGCATGGGTTCACGTGGCCCCAGATGCAGCCTGGGACCAGGCTCCCTGGTGCTGATGGTGGGACAGTCACCCTGGGG
GTTGACCGCCGGACTGGGCGTCCCCAGGGTTGACTATAGGACCAGGTGTCCAGGTGCCCTGCAAGTAGAGGGGCTCTCAG
AGGCGTCTGGCTGGCATGGGTGGACGTGGCCCCGGGCATGGCCTTCAGCGTGTGCTGCCGTGGGTGCCCTGAGCCCTCAC
TGAGTCGGTGGGGGCTTGTGGCTTCCCGTGAGCTTCCCCCTAGTCTGTTGTCTGGCTGAGCAAGCCTCCTGAGGGGCTCT
CTATTGCAG

Intron 4 (SEQ ID NO 7)

GTGGCTGTGCTTTGGTTTAACTTCCTTTTTAAACAGAAGTGCGTTTGAGCCCCACATTTGGTATCAGCTTAGATGAAGGG
CCCGGAGGAGGGGCCACGGGACACAGCCAGGGCCATGGCACGGCGCCAACCCATTTGTGCGCACAGTGAGGTGGCCGAGG
TGCCGGTGCCTCCAGAAAAGCAGCGTGGGGGTGTAGGGGGAGCTCCTGGGGCAGGGACAGGCTCTGAGGACCACAAGAAG
CAGCCGGGCCAGGGCCTGGATGCAGCACGGCCCGAGGTCCTGGATCCGTGTCCTGCTGTGGTGCGCAGCCTCCGTGCGCT
TCCGCTTACGGGGCCCGGGGACCAGGCCACGACTGCCAGGAGCCCACCGGGCTCTGAGGATCCTGGACCTTGCCCCACGG
CTCCTGCACCCCACCCCTGTGGCTGCGGTGGCTGCGGTGACCCCGTCATCTGAGGAGAGTGTGGGGTGAGGTGGACAGAG
GTGTGGCATGAGGATCCCGTGTGCAACACACATGCGGCCAGGAACCCGTTTCAAACAGGGTCTGAGGAAGCTGGGAGGGG
TTCTAGGTCCCGGGTCTGGGTGGCTGGGGACACTGGGGAGGGGCTGCTTCTCCCCTGGGTCCCTATGGTGGGGTGGGCAC
TTGGCCGGATCCACTTTCCTGACTGTCTCCCATGCTGTCCCCGCCAG

Intron 5 (SEQ ID NO 8)

GTGGGTGCCGGGGACCCCCGTGAGCAGCCCTGCTGGACCTTGGGAGTGGCTGCCTGATTGGCACCTCATGTTGGGTGGAG
GAGGTACTCCTGGGTGGGCCGCAGGGAGTGCAGGTGACCCTGTCACTGTTGAGGACACACCTGGCACCTAGGGTGGAGGC
CTTCAGCCTTTCCTGCAGCACATGGGGCCGACTGTGCACCCTGACTGCCCGGGCTCCTATTCCCAAGGAGGGTCCCACTG
GATTCCAGTTTCCGTCAGAGAAGGAACCGCAACGGCTCAGCCACCAGGCCCCGGTGCCTTGCACCCCAGTCCTGAGCCAG
GGGTCTCCTGTCCTGAGGCTCAGAGAGGGGACACAGCCCGCCCTGCCCTTGGGGTCTGGAGTGGTGGGGGTCAGAGAGAG

AGTGGGGGACACCGCCAGGCCAGGCCCTGAGGGCAGAGGTGATGTCTGAGTTTCTGCGTGGCCACTGTCAGTCTCCTCGC
CTCCACTCACACAG

5'-Bereich Intron 6 (SEQ ID NO 9)

GTAAGGTTCACGTGTGATAGTCGTGTCCAGGATGTGTGTCTCTGGGATATGAATGTGTCTAGAATGCAGTCGTGTCTGTG
ATGCGTTTCTGTGGTGGAGGTACTTCCATGATTTACACATCTGTGATATGCGTGTGTGGCACGTGTGTGTCGTGGTGCAT
GTATCTGTGGCGTGCATATTTGTGGTGTGTGTGTGTGTGGCACGTGTGTGTCCATGGTGTGTGTGCCTGTGGTGTGCATG
TGTGTGTGTCTGTGACACGTGCATGTTCATGCTGTGTGCTGCATGTCTGTGATGTGCCTATTTGTGGTGTGTGTGTGCAT
GTGTCCGTGACATATGCGTGTCTATGGCATGGGTGTGTGTGGCCCCTTGGCCTTACTCCTTCCTCCTCCAGGCATGGTCC
GCACCATTGTCCTCACGCTCTCGGGTGCTGGTTTGGGGAGCTCCACATTCAGGGTCCTCACTTCTAGCATGGGTGCCCCT
GTCCTGTCACAGGGCTGGGCCTTGGAGACTGTAAGCCAGGTTTGAGAGGAGAGTAGGGATGCTGGTGGTACCTTCCTGGA
CCCCTGGCACCCCCAGGACCCCAGTCTGGCCTATGCCGGCTCCATGAGATATAGGAAGGCTGATTCAGGCCTCGCTCCCC
GGGACACACTCCTCCCAGAGCGGCCGGGGGCCTTGGGGCTCGGCAGGGGTGAAAGGGGCCCTGGGCTTGGGTTCCCACCC
AGTGGTCATGAGCACGCTGGAGGGGTAAGCCCTCAAAGTCGTGCCAGGCCGGGGTGCAGAGGTGAAGAAGTATCCCTGGA
GCTTCGGTCTGGGGAGAGGCACATGTGGAAACCCACAAGGACCTCTTTCTCTGACTTCTTGAGCT

3'-Bereich Intron 6 (SEQ ID NO 10)

TGTGGGATTGGTTTTCATGTGTGGGATAGGTGGGGATCTGTGGGATTGGTTTTTATGAGTGGGGTAACACAGAGTTCAAG
GCGAGCTTTCTTCCTGTAGTGGGTCTGCAGGTGCTCCAACAGCTTTATTGAGGAGACCATATCTTCCTTTGAACTATGGT
CGGGTTTATAGTAAGTCAGGGGTGTGGAGGCCTCCCCTGGGCTCCCTGTTCTGTTTCTTCCACTCTGGGGTCGTGTGGTG
CCTGCTGTGGTGTGTGGCCGGTGGGCAGGGCTTCCAGGCCTCCTTGTGTTCATTGGCCTGGATGTGGCCCTGGCTACGCT
CCGTCCTTGGAATTCCCCTGCGAGTTGGAGGCTTTCTTTCTTTCTTTTTTTCTTTCTTTTTTTTTTTTTTTGATAACAGA
GTCTCGCTCTTTTTTGCCCAGGCTGGAGTGGTTTGGCGTGATCTTGGCTCACTGCAACCTGTGCTTCCTGAGTTCAAGCA
ATTCTCTTGCCTCAGCCTCCCAAGTAGCTGGAATTATAGGCGCCCACCACCATGCTGACTAATTTTTGTAATTTTAGTAG
AGACGAGGTTTCTCCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCTCCCACCTCGGCCTCCCAAAGT
GCTGGGATGACAGGTGTGAACCGCCGCGCCCGGCCGAGACTCGCTTCCTGCAGCTTCCGTGAGATCTGCAGCGATAGCTG
CCTGCAGCCTTGGTGCTGACAACCTCCGTTTTCCTTCTCCAGGTCTCGCTAGGGGTCTTTCCATTTCATGACTCTCTTCA
CAGAAGAGTTTCACGTGTGCTGATTTCCCGGCTGTTTCCTGCGTAATTGGTGTCTGCTGTTTATCGATGGCCTCCTTCCA
TTTCCTTTAGGCTTTGTTTATTGTTGTTTTTCCGGCTCCTTGAAGGAAAAGTTTCGATTATGGATGTTTGAACTTTCTTT
TCTAAACAAGCATCTGAAGTTGCCGTTTTCCCTCTAAAGCAGGGATCCCGAGGCCCCTGGCTGTGGAGTGGCACCGGTCT
GGGGCCTGTTAGGAACCCGGCGCACAGCGGGAGGCTAGGTGGGGTGTGGGGAGCCAGCGTTCCCGCCTGAGCCCCGCCCC
TCTCAGATCAGCAGTGGCATGCGGTGCTCAGAGGCGCACACACCCTACTGAGAACTGTGCGTGAGAGGGGTCTAGATTCT
GTGCTCCTTATGGGAATCTAATGCCTGATGATCTGAGGTGGAACCGTTTGCTCCCAAAACCATCCCCTTCCCCACTGCTG
TCCTGTGGAAAAATCGTCTTCCACGAAACCAGTCCCTGGTACCACAATGGTTGGGGACCCTGTGCTAAAGACCTGCTTCA
GCAGCCTCTCGTCAGTGTTGATATATTGGCTTTTCTGTGTTGAGTCCAGAATAATTACGGATTTCTGTGATGCTTTCCGC
CGACCTCAGACCCATGGGCTATTTGTGGGCGTGTTGCCTGCTCCTGGGTTGGGAAGGGTGCAGGCCCCATGTACCTTCCT
GTTACTGCCTTCCAGGTTGGTTCTCAGGGTTGAATCGTACTCGATGTGGTTTTAGCCCACGGCCCTGCCGCCAGCTCCTG
GGGGCTGGGGAACATGCTGAAGCACAGAGTCACCGTGCGCGTCTTTTGATGCCTCACAAGCTCGAGGCCTCCTGTGTCCG
TGTTAGTGTGTGTCACGTGCCTGCTCACATCCTGTCTTGGGGACGCAGGGGCTTAGCAGGTCCCGTAGTAAATGACAAGC
GTCCTGGGGGAGTCTGCAGAATAGGAGGTGGGGGTGCCGGTCTCTCTCCCGCGTCTTCAGACTCTTCTCCTGCCTGTGCT

GTGGCTGCACCTGCATCCCTGCAATCCCTCCAGCACTGGGCTGGAGAGGCCCGGGAGCTCGAGTGCCACTTGTGCCACGT
GACTGTGGATGGCAGTCGGTCACGGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTTGGTCACAGGGGTCTGATGTGTG
GTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGG
ATGGCGGTCGTGGGGTCTGATGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTGACTGTGGATGGCGGTCGTG
GGGTCTGATGTGGTGACTGTGGATGGCAGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATG
TGGTGACTGTGGATGGCAGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACT
GTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGG
CGGTCGTGGGGTCTGATGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGTGATCGGTCA
CAGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGTGATCGGTCACAG
GGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTTGGTCCCGGGGG
TCTGATGTGTGGTGACTGTGGATGGCGATCGGTCACAGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCT
GATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGGT
GACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGAT
GGCGGTTGGTCCCGGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTGACTGTGGATGGCAG
TCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGG
TCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGT
GGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGTGATCGGTCACAGGGGTCTGATGTGTGGT
GACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTGACTGTGGAT
GGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTAGGGTCTGATGTGTGGTGACTGTGGATGGCAGTCG
GTCACAGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGG
GGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGAT
GTGGTGACTGTGGATGGTGATCGGTCACAGGGGTCTGATGTGTGGTAGCTGCAGGTGGAGTCCCAGGTGTGTCTGTAGCT
ACTTTGCGTCCTCGGCCCCCCGGCCCCCGTTTCCCAAACAGAAGCTTCCCAGGCGCTCTCTGGGCTTCATCCCGCCATCG
GGCTTGGCCGCAGGTCCACACGTCCTGATCGGAAGAAACAAGTGCCCAGCTCTGGCCGGGGCAGGCCACATTTGTGGCTC
ATGCCCTCTCCTCTGCCGGCAG

Intron 7 (SEQ ID NO 11)

GTCTGGGCACTGCCCTGCAGGGTTGGGCACGGACTCCCAGCAGTGGGTCCTCCCCTGGGCAATCACTGGGCTCATGACCG
GACAGACTGTTGGCCCTGGGGGGCAGTGGGGGGAATGAGCTGTGATGGGGGCATGATGAGCTGTGTGCCTTGGCGAAATC
TGAGCTGGGCCATGCCAGGCTGCGACAGCTGCTGCATTCAGGCACCTGCTCACGTTTGACTGCGCGGCCTCTCTCCAGTT
CCGCAGTGCCTTTGTTCATGATTTGCTAAATGTCTTCTCTGCCAGTTTTGATCTTGAGGCCAAAGGAAAGGTGTCCCCCT
CCTTTAGGAGGGCAGGCCATGTTTGAGCCGTGTCCTGCCCAGCTGGCCCCTCAGTGCTGGGTCTGAGGCCAAAGGAAACG
TGTCCCCCTTCTTAGGAGGACGGGCCGTGTTTGAGCCACGCCCCGCTGAGCGGGCCTCTCAGTGCTGGGTCTGTCCACGT
GGCCCTGTGGCCCTTTGCAGATGTGGTCTGTCCACGTGGCCCTGTGGCTCTTTGCAGATGCCTGTTAGCACTTGCTCGGC
TCTAGGGGACAGTCGTGTCCACCGCATGAGGCTCAGAGACCTCTGGGCGAATTTCCTTGGCTCCCAGGGTGGGGGTGGAG
GTGGCCTGGGCTGCTGGGACCCAGACCCTGTGCCCGGCAGCTGGGCAGCAACTCCTGGATCACATATGCCATCCGGGCCA
CGGTGGGCTGTGTGGGTGTGAGCCCAGCTGGACCCACAGGTGGCCCAGAGGAGACGTTCTGTGTCACACACTCTGCCTAA
GCCCATGTGTGTCTGCAGAGACTCGGCCCGGCCAGCCCACGATGGCCCTGCATTCCAGCCCAGCCCCGCACTTCATCACA
AACACTGACCCCAAAAGGGACGGAGGGTCTTGGCCACGTGGTCCTGCCTGTCTCAGCACCCACCGGCTCACTCCCATGTG
TCTCCCGTCTGCTTTCGCAG

Intron 8 (SEQ ID NO 12)

GTGAGTCAGGTGGCCAGGTGCCATTGCCCTGCGGGTGGCTGGGCGGGCTGGCAGGGCTTCTGCTCACCTCTCTCCTGCCC
CTTCCCCACTGNCCTTCTGCCCGGGGCCACCAGAGTCTCCTTTTCTGGCCCCCGCCCCCTCCGGCTCCTGGGCTGCAGGC
TCCCGAGGCCCCGGAAACATGGCTCGGCTTGCGGCAGCCGGAGCGGAGCAGGTGCCACACGAGGCCTGGAAATGGCAAGC
GGGGTGTGGAGTTGCTCCTGCGTGGAGGACGAGGGGCGGGGGGTGTGTCTGGGTCAGGTGTGCGCCGAGCGTTTGAGCCT
GCAGCTTGTCAGCTCCAAGTTACTACTGACGCTGGACACCCGGCTCTCACACGCTTGTATCTCTCTCTCCCGATACAAAA
GGATTTTATCCGATTCTCATTCCTGTCCCTGTCGTGTGACCCCCGCGAGGGCGCGGGCTCTTCTCTCTGTGACTAGATTT
CCCATCTGGAAAGTGCGGGGTTGACCGTGTAGTTTGCTCCTCTCGGGGGGCCTGTGGTGGCCATGGGGCAGGCGGCCTGG
GAGAGCTGCCGTCACACAGCCACTGGGTGAGCCACACTCACGGTGGTAGAGCCACAGTGCCTGGTGCCACATCACGTCCT
CTGGATTTTAAGTAAAACCACACACCTCCCGGCAGGCATCTGCCTGCGACCCTGTGTGTGCCTGGGGAGAGTGGTAGCAC
GGAGGAAATTCGTGCACACTCAAGGTCATCAGCAAGGTCATCCGCAGTCAGGTGGAACGTGGAGGCCTCTCTCTGGGATC
GTCTCCAGCGGATAAAGGACTGTGCACAGCTTCGGAAGCTTTTATTTAAAAATATAACTATTAATTATTGCATTATAAGT
AATCACTAATGGTATCAGCAATTATAATATTTATTAAAGTATAATTAGAAATATTAAGTAGTACACACGTTCTGGAAAAA
CACAAATTGCACATGGCAGCAGAGTGAATTTTGGCCGAGGGACACGTGTGCACATGTGTGTAAGCGGCCCCCAGGCCCAC
AGAATTCGCTGACAAAGTCACCTCCCCAGAGAAGCCACCACGGGCCTCCTTCGTGGTCGTGAATTTTATTAAGATGGATC
AAGTCACGTACCGTCCACGTGTGGCAGGGCTTTGGGGAATGTGAGGTGATGACTGCGTCCTCATGCCCTGACAGACAGGA
GGTGACTGTGTCTGTCCTGTCCCTAGGACACGGACAGGCCCGAAGCTCTAGTCCCCATCGTGGTCCAGTTTGGCCTCTGA
ATAAAAACGTCTTCAAAACCTGTTGCCCCAAAAACTAAGAACAGAGAGAGTTTCCCATCCCATGTGCTCACAGGGGCGTA
TCTGCTTGCGTTGACTCGCTGGGCTGGCCGGACTCCTAGAGTTGGTGCGTGTGCTTCTGTGCAAAAAGTGCAGTCCTCTT
GCCCATCACTGTGATATCTGCACCAGCAAGGAAAGCCTCTTTTCTTTTCTTTCTTTTTTTTTTTTTGAGACGGAACGTCA
CTGTTGTCTGCCTGGGCTTGAGTGCAGTGGCGCGATCTCAACTCACTGCAACCTCCGCCTCCCGGGTTCCAGCATTTCTC
CTGCCTCAGCCTCCCGAGCAGCTGAGATTACAGGCACCCACCCCCTGCGCCTGGCTAATTTTTGTATTTTTAGTAGAGAG
GGGTTTTTGCCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCACCCACCTCGGCCTCCCAAAGTGCTG
GGATTACAGGTGTGAGCCATCACGCCCAGCCGGAAAGCCTCTTTTTAAGGTGACCACCTATAGCGCTTCCCGAAAATAAC
AGGTCTTGTTTTTGCAGTAGGCTGCAAGCGTCTCTTAGCAACAGGAGTGGCGTCCTGTGGGCTCTGGGGATGGCTGAGGG
TCGCGTGGCAGCCATGCCTTCTGTGTGCACCTTTAGGTTCCACGGGGCTATTCTGCTCTCACTGTTTGTCTGAAAACGCA
CCCTTGGCATCCTTGTTTGGAGAGTTTCTGCTTCTCGTTGGTCATGCTGAAACTAGGGGCAAGGTTGTATCCGTTGGCGC
GCAGCGGCTACATGTAGGGTCATGAGTCTTTCACCGTGGACAAATTCCTTGAAAAAAAAAAAAAGGAGTCCGGTTAAGCAT
TCATTCCGGGTCAAGTGTCTGGTTCTGTGAATAAACTCTAAGATTTAAGAAACCTTAATGAAAGAAAACCTTGATGATTC
AGAGCAAGGATGTGGTCACACCTGTGGCTGGATCTGTTTCAGCCGCCCCAGTGCATGGTGAGAGTGGGGAGCAGGGATTG
TTTGTTCAGAGGTCTCATCTGGTATGTTTCTGAGGTGTTTGCCGGCTGAATGGTAGACGTGTCGTTTGTGTGTATGAGGT
TCTGTGTCTGTGTGTGGCTCGGTTTGAGTGTACGCATGTCCAGCACATGCCCTGCCCGTCTCTCACCTGTGTCTTCCCGC
CCCAG

Intron 9 (SEQ ID NO 13)

GTGAGGCCTCCTCTTCCCCAGGGGGGCTTGGGTGGGGGTTGATTTGCTTTTGATGCATTCAGTGTTAATATTCCTGGTGC
TCTGGAGACCATGACTGCTCTGTCTTGAGGAACCAGACAAGGTTGCAGCCCCTTCTTGGTATGAAGCCGCACGGGAGGGG
TTGCACAGCCTGAGGACTGCGGGCTCCACGCAGGCTCTGTCCAGCGGCCATGTCCAGAGGCCTCAGGGCTCAGCAGGCGG
GAGGGCCGCTGCCCTGCATGATGAGCATGTGAATTCAACACCGAGGAAGCACACCAGCTTCTGTCACGTCACCCAGGTTC

CGTTAGGGTCCTTGGGGAGATGGGGCTGGTGCAGCCTGAGGCCCCACATCTCCCAGCAGGCCCTCGACAGGTGGCCTGGA
CTGGGCGCCTCTTCAGCCCATTGCCCATCCCACTTGCATGGGGTCTACACCCAAGGACGCACACACCTAAATATCGTGCC
AACCTAATGTGGTTCAACTCAGCTGGCTTTTATTGACAGCAGTTACTTTTTTTTTTTTAATACTTTAAGTTCTAGGGTAC
ATGTGCACGACGTGCAGGTTAGTTACATATGTATACATGTGCCATGTTGGTGTGCTGCACCCATTAACTCATCATTTACA
TTAGGTATATCTCCTAATGCTATCCCTCCCCACTCCCCCCATCCCATGACAGGCCCTGGTGTGTGATGTTCCCCACCCTG
TGTCCAAGTGTTCTCATTGTTCAGTTCCCACCTGTGAGTGAGAACATGTGGTGTTTGGTTTTCTTTCCTTGCAATAGTTT
GCTCAGAGTGATGGTTTCCAGCTTCGTCCATGTCCCTACAAAGGACATGAACTCATCCTTTTTTATGACTGCATAGTATT
CCGTGGTGTATATGTGCCACATTTTCTTAATCCAGTCTATCATCGATGGACATTTGGGTTGGTTGCAAGTCTTTGCTACT
GTGAATAGTGCCGCAATAAACATACGTGTGCATGTGTCTTTATAGCAGCATGATTTATAATCCTTTGGGTATATACCCAG
TAATGGGATGGCTGGGTCAAATGGTATTTCTAGTTCTAGATCCTTGAGGAATCACCACACTGTCTTCCACAATGGTTGAA
CTAGTTTACACTCCCACCAACAGTGTAAAAGTGTTCTGGTGCTGGAGAGGATGTGGACAGCAGTTATTTTTTTATGAAAA
TAGTATCACTGAACAAGCAGACAGTTAGTGAAGGATGCGTCAGGAAGCCTGCAGGCCACACAGCCATTTCTCTCGAAGAC
TCCGGGTTTTTCCTGTGCATCTTTTGAAACTCTAGCTCCAATTATAGCATGTACAGTGGATCAAGGTTCTTCTTCATTAA
GGTTCAAGTTCTAGATTGAAATAAGTTTATGTAACAGAAACAAAAATTTCTTGTACACACAACTTGCTCTGGGATTTGGA
GGAAAGTGTCCTCGAGCTGGCGGCACACTGGTCAGCCCTCTGGGACAGGATACCTCTGGCCCATGGTCATGGGGCGCTGG
GCTTGGGCCTGAGGGTCACACAGTGCACCATGCCCAGCTTCCTGTGGATAGGATCTGGGTCTCGGATCATGCTGAGGACC
ACAGCTGCCATGCTGGTAAAGGGCACCACGTGGCTCAGAGGGGGCGAGGTTCCCAGCCCCAGCTTTCTTACCGTCTTCAG
TTATTTTTCCCTAAGAGTCTGAGAAGTGGGGCCGCGCCTGATGGCCTTCGTTCGTCTTCAGCTGGCACAGAATTGCACAA
GCTGATGGTAAACACTGAGTACTTATAATGAATGAGGAATTGCTGTAGCAGTTAACTGTAGAGAGCTCGTCTGTTGGAAA
GAAATTTAAGTTTTTCATTTAACCGCTTTGGAGAATGTTACTTTATTTATGGCTGTGTAAATTGTTTGACATTCAGTCCC
TCGTAGACAGATACTACGTAAAAAGTGTAAAGTTAACCTTGCTGTGTATTTTCCCTTATTTTAG

Intron 10 (SEQ ID NO 14)

GTGAGGCCCGTGCCGTGTGTCTGTGGGGACCTCCACAGCCTGTGGGCTTTGCAGTTGAGCCCCCCGTGTCCTGCCCCTGG
CACCGCAGCGTTGTCTCTGCCAAGTCCTCTCTCTCTGCCGGTGCTGGATCCGCAAGAGCAGAGGCGCTTGGCCGTGCACC
CAGGCCTGGGGGCGCAGGGGCACCTTCGGGAGGGAGTGGGTACCGTGCAGGCCCTGGTCCTGCAGAGACGCACCCAGGTT
ACACACGTGGTGAGTGCAGGCGGTGACCTGGCTCCTGCTGCTCTTTGGAAAGTCAAGAGTGGCGGCTCCTGGGGCCCCAG
TGAGACCCCCAGGAGCTGTGCACAGGGCCTGCAGGGCCGAGGCGGCAGCCTCCTCCCCAGGGTGCACCTGAGCCTGCGGA
GAGCAGGAGCTGCTGAGTGAGCTGGCCCACAGCGTTCGCTGCGGTCACGTTCCTGCGTGGGGTTGTTTGGGATCGGTGGG
AGAATTTGGATTTGCTGAGTGCTGCTGTCTTGAACCACGGAGATGGCTAGGAGTGGGTTTCAGAGTTGATTTTTGTGAAT
CAAACTAAAATCAGGCACAGGGGACCTGGCCTCAGCACAGGGGATTGTCCAATGTGGTCCCCCTCAAGGGCGCCCCACAG
AGCCGGTGGGCTTGTTTTAAAGTGCGATTTGACGAGGGACGAGAAACCTTGAAAGCTGTAAAGGGAACCCTCAGAAAATG
TGGCCGCCAGGGGTGGTTTCAGGTGCTTTGCTGGGCTGTGTTTGTGAAAACCCATTTGGACCCGCCCTCCAAGTCCACCC
TCCAGGTCCACCCTCCAGGGCCGCCCTGGGCTGGGGGTATGCCTGGCGTTCCTTGTGCCGCAGCCCGGAGCACAGCAGGC
TGTGCACATTTAAATCCACTAAGATTCACTCGGGGGGAGCCCAGGTCCCAAGCAACTGAGGGCTCAGGAGTCCTGAGGCT
GCTGAGGGGACAGAGCAGACGGGGAACGCTGCTTCTGTGTGGCAAGTTCCTGAGGGTGCTGGCCAGGGAGGTGGCTCAGA
GTGTATGTTGGGGTCCCACCGGGGGCAGAACTCTGTCTCTGATGAGTCGGCAGCCATGTAACAGGAAGGGGTGGCCACAG
GGAGCTGGGAATGCACCAGGGGAGCTGCGCAGCTGGCCGAGGTCCCAGGGCCAGGCCACAGGAAGGGCAGGGGGACGCCC
GGGGCCACAGCAGAGGCCGCAGGAAGGGAAGGGGATGCCCAGGCCAGAGCAGAGGCTACCGGGCACAGGGGGGCTCCCTG
AGCTGGGTGAGCGAGGCTCATGACTCGGCGAGGGAACCTCCTTGACGTGAAGCTGACGACTGGTGTTGCCCAGCTCACAG

CCCAGCCAGGTCCCGCGCCTGAGCAGGAACTCAGAACCCTCCCCTTTGTCTAAAGCACAGCAGATGCCTTCAGGGCATCT
AGGAGAAAACAGGCAAAGTCGTTGAGAAACGTCTTAAAAGAAGGTGGGATGGTGGCAATTTCTTGTCCAGATTTTAGTCT
GCCCCGGACCACAGATGAGTCTATAACGGGATTGTGGTGTTGCCATGGGGACACATGAGATGGACCATCACAGAGGCCAC
TGGGGCTGCACCTCCCATCTGAGTCCTGGCTGTCCCGGGTCCAGGCCAGGTTCTTGCATGCTCACCTACCTGTCCTGCCC
GGGAGACAGGGAAAGCACCCCGAAGTCTGGAGCAGGGCTGGGTCCAGGCTCCTCAGAGCTCCTGCCAGGCCCAGCACCCT
GCTCCAAATCACCACTTCTCTGGGGTTTTCCAAAGCATTTAACAAGGGTGTCAGGTTACCTCCTGGGTGACGGCCCCGCA
TCCTGGGGCTGACATTGCCCCTCTGCCTTAG

Intron 11 (SEQ ID NO 15)

GTGAGCGCACCTGGCCGGAAGTGGAGCCTGTGCCCGGCTGGGGCAGGTGCTGCTGCAGGGCCGTTGCGTCCACCTCTGCT
TCCGTGTGGGGCAGGCGACTGCCAATCCCAAAGGGTCAGAGGCCACAGGGTGCCCCTCGTCCCATCTGGGGCTGAGCAGA
AATGCATCTTTCTGTGGGAGTGAGGGTGCTCACAACGGGAGCAGTTTTCTGTGCTATTTTGGTAAAAGGAAATGGTGCAC
CAGACCTGGGTGCACTGAGGTGTCTTCAGAAAGCAGTCTGGATCCGAACCCAAGACGCCCGGGCCCTGCTGGGCGTGAGT
CTCTCAAACCCGAACACAGGGGCCCTGCTGGGCATGAGTCCCTCTGAACCCGAGACCCTGGGGCCCTGCTGGGCGTGAGT
CTCTCCGAACCCAGAGACTTCAGGGCCCTTTTGGGCGTGAGTCTCTCCGCTGTGAGCCCCACACTCCAAGGCTCATCCAC
AGTCTACAGGATGCCATGAGTTCATGATCACGTGTGACCCATCAGGGGACAGGGCCATGGTGTGGGGGGGGTCTCTACAA
AATTCTGGGGTCTTGTTTCCCCAGAGCCCGAGAGCTCAAGGCCCCGTCTCAGGCTCAGACACAAATGAATTGAAGATGGA
CACAGATGCAGAAATCTGTGCTGTTTCTTTTATGAATAAAAAGTATCAACATTCCAGGCAGGGCAAGGTGGCTCACACCT
ATAATCCCAGCACTTTGGGAGGCCGAGGTGGGTGGATCACTTGAGGCCAGGAGTTTGAGGCCAACCTAACCAACATAGTG
AAATTCCATTTCTACTTAAAAAATACAAAAATTAGCCTGGCCTGGTGGCACACGCCTGTAGTCCCCGCTATGCGGGAGGC
TGAGGCAGGAGAATCATTTGAACCCAGGAGGCAGAGGTTGCAGTGAGCCGAGATCACACCACTGCACTCCAGCCTGGGCA
ACAGAGTGAGACTTCATCTTAAAAAAAAAAAAAAAAAGTATCAGCATTCCAAAACCATAGTGGACAGGTGTTTTTTTATTC
TGTCCTTCGATAATATTTACTGGTGCTGTGCTAGAGGCCGGAACTGGGGGTGCCTTCCTCTGAAAGGCACACCTTCATGG
GAAGAGAAATAAGTGGTGAATGGTTGTTAAACCAGAGGTTTAAACTGGGGTCCTGTCGTTCTGAGTTAACAGTCCAGATC
TGGACTTTGCCTCTTTCCAGAATGCTCCCTGGGGTTTGCTTCATGGGGGAGCAGCAGGTGTGGACACCCTCGTGATGGGG
GAGCAGCAGGTGCAGACGCCCTCATGATGGGGGAGTGGCAGGTGCAGACACCCTTGTGCATGGTGCCCAGCATGTCCCTG
TTGCAGCTCCCTCCCCACAAGGATGCCGGTCTCCTGTGCTCCCCACAGTCCCTGCTTCCCTCTCACAGCCTTACCTGGTC
CTGGCCTCCACTGGCTTTGTCTGCATGATTTCCACATTTCCTGGGCTCCCAGCACCTCTTCGCCTCTCCCAGGCACCTCT
GCAGTGCTGGCCATACCAGTCAGCTGTGAACTGTCCACTGCTTATTTTGCTCCCCATGAAATGTATTTTTTAGGACAGGC
ACCCCTGGTTCCAGCCTCTGGCACAGCATCAGTGAATGTTATTGAAGGACAAAGGACAGACAAACAAATCAGGAAAATGG
GTTCTCTCTAAACACATTGCAAAGCCACAGAGGCTAGTGCAGGATGGGTGGGCATCAGGTCATCAGATGTGGGTCCAATG
CCAGAATATTCTGTGCTCCCAAAGGCCACTTGGTCAGAGTGTGTGCTTGCAGAGGTGGCTCTAAAAGCTCAGCAGTGGAG
GCAGTGGTTCGCCATACTCAGGGTGAACTCACATCCTCTGTGTCTGAAGTATACAGCAGAGGCTTGAAGGGCATCTGGGA
GAAGAAAACAGGCAAAATGATTAAGAAAAGTGAAAAAGGAAAAGTGGTAAGATGGGAATTTTCTTGTCCAGATTTTAGTC
TCCCAAACCACAGCTCAGATGGTAGAATGTGGTCAGAACTGATGGACAGAACAATAGAACAAAACGGAAGCCCTATCTCT
CAGAAACGTGTGTTAATGTGGTATGTGGCACAGCTGATGGAAAAGAGAGTGTGTGTGTAATTTTTTTTTCTGAGAAAACT
GACTGGAAGCAAATAAGTTGTGTCTTTACAGCATATACCAGAGCAGATTCTAGGTAGAAGAGGAGACACATGCAAACAAC
ACCAGCAACAGAAATAAAACAAAAGACTCAAAGGGAAGGGAGGTGAACGTTCCCTGGTTTGGTGTTGGGGAAGGACACAC
AGGGAGGCGGATGAAACCAGTGAGGCAACGGGCATTGCTTTCACTGCAGAGAAACTCAGCTTGCCTGAGCCACAGTGAAA
ATGGCCATTCCCTGGAGCGTTTGTGCACGTGATTTATTTAAGGCGCCCTGTGAGGTCCTGCACATTCATCCTCTCACTTT

GTTCTCCTAACCACCTGAGAGGTAGAGGAGGAAAGGCTCCAGGGGAGCAGCCGCCCTTGGTCACCCAGCTGGCAAAGGGC
ATGCATGATTGCAGCCTGGCCTCCTGCTCCGGGGCCCTTGCTCTGCCCGAGGACCCCACACAAGTCAGACCCATAGGCTC
AGGGTGAGCCGGAGCCCAAGGTCGTGTTGGGGATGGCTGTGAAAGAAGAAATGGACGTCTGATGCACACTTGGGAAGGTC
CTACCAGCAGCGTCAAAGAAATGCATGTGAAACTGACAGCGAGACCCATCCCTCAAAGAAACGCACGTGAAACTGATGGC
GAGACCTGTCCCCATCCCTCATGCTGGCTCCTTTTCTGGGCTTGCCAAGAGCCAGCATCAGGTTGAGGCAAGCTGGAAAG
ACTTTTCTGGAAAGCAGCTTGTTTGCATGGAAGTCCTCACAATGTCCTGTGTCTTCCCAGTAATTCCACTTCTGAAGTGA
CCAGACATTATCACGGGTCTTATTTACCATTTCCAGTGTTCCAGGCAGGGGGACTTGCCACAGCAAGTCACGAACCTGCC
CAAATACAGGGCTAAGGAGATATTATGCATCACAAAACTTGCTCTGCCATTAAACATTTTTCAAAGAATTTTTGAAGAAT
GTTTAATGGCACAAAACGTTTATTTCAATGTAGCAGTGTTCAAAGCTGGATGTAAAAGAACACACCCCAGGAGCCTGCCG
TGAATGTCATGTGTGTTCATCTTTGGACATGGACATACATGGGCAGTGAGTGGTGGTGAGGCCCTGGAGGACATCGGTGG
GATGCCTCCATCCTGCCCCTCTGGAGACACCATGTGTGCCACGTGCACTCACTGGAGCCCTGTTTAGCTGGTGCCACCTG
GCTCTTCCATCCCTGAGATTCAAACACAGTGAGATTCCCCACGCCCAACTCAGTGTTCTCCCACAAAAAACCTGAGTCAC
ACCTGTGTTCACTCGAGGGACGCCCGGGAGCCAGGGCTCCACAGTTTATTATGTGTTTTTGGCTGAGTTATGTGCAGATC
TCATCAGGGCAGATGATGAGTGCACAAACACGGCCGTGCGAGGTTTGGATACACTCAACATCACTAGCCAGGTCCTGGTG
GAGTTTGGTCATGCAGAGTCTGGATGGCATGTAGCATTTGGAGTCCATGGAGTGAGCACCCAGCCCCCTCGGGCTGCAGC
GCATGCCCCAGGCAGGACAAGGAAGCGGGAGGAAGGCAGGAGGCTCTTTGGAGCAAGCTTTGCAGGAGGGGGCTGGGTGT
GGGGCAGGCACCTGTGTCTGACATTCCCCCCTGTGTCTCAG

Intron 12 (SEQ ID NO 16)

GTGAGCAGGCTGATGGTCAGCACAGAGTTCAGAGTTCAGGAGGTGTGTGCGCAAGTATGTGTGTGTGTGTGCGCGCGT
GCCTGCAAGGCTGATGGTGACTGGCTGCACGTAAGAGTGCACATGTACGCATATACACGTGAGCACATACATGTGTGCAT
GTGTGTACATGAAGGCATGGCAGTGTGTGCACAGGTGTGCAAGGGCACAAGTGTGTGCACATGCGAATGCACACCTGACA
TGCATGTGTGTTCGTGCACAGTCGTGTGGGCATTCACGTGAGGTGCATGCGTGTGGGTGTGCAGTGTGAGTAGCATGTGT
GCACATAACATGTATTGAGGGGTCCTCGTGTTCACCCCGCTAGGTCCTCAGCACCAGTGCCACTCCTTACAGGATGAGAC
GGGGTCCCAGGCCTTGGTGGGCTGAGGCTCTGAAGCTGCAGCCCTGAGGGCATTGTCCCATCTGGGCATCCGCGTCCACT
CCCTCTCCTGTGGGCTTCTGTGTCCACTCCCCCTCTCCTGTGGGCATTTACATCCACTCCACTCCCTCTCTCCTGTGGGC
ATCCGCGTCCACTCCCCCTCTCTGTGGGCATCTGCGTCCACCTCCCCTCTCTGTGGGCATTTGCGTCCACTCCCTCTCCT
GGTTCCTTCCTGTCTTGGCCGAGCCTCGGGGGCAGGCAGATGACACAGAGTCTTGACTCGCCCAGGGTGGTTCGCAGCTG
CCGGGTGAGGGCCAGGCCGGATTTCACTGGGAAGAGGGATAGTTTCTTGTCAAAATGTTCCTCTTTCTTGTTCCATCTGA
ATGGATGATAAAGCAAAAAGTAAAAACTTAAAATCCCAGAGAGGTTTCTACCGTTTCTCACTCTTTCTTGGCGACTCTAG

Intron 13 (SEQ ID NO 17)

GTGAGCCGCCACCAAGGGGTGCAGGCCCAGCCTCCAGGGACCCTCCGCGCTCTGCTCACCTCTGACCCGGGGCTTCACCT
TGGAACTCCTGGGTTTTAGGGGCAAGGAATGTCTTACGTTTTCAGTGGTGCTGCTGCCTGTGCACAGTTCTGTTCGCGTG
GCTCTGTGCAAAGCACCTGTTCTCCATCTCTGGGTAGTGGTAGGAGCCGGTGTGGCCCCAGGTGTCCCCACTGTGCCTGT
GCACTGGCCGTGGGACGTCATGGAGGCCATCCCAGGGCAGCAGGGGCATGGGGTAAAGAGATGTTTATGGGGAGTCTTAG
CAGAGGAGGCTGGGAAGGTGTCTGAACAGTAGATGGGAGATCAGATGCCCGGAGGATTTGGGGTCTCAGCAAAGAGGGCC
GAGGTGGGTGCAGGTGAGGGTCGCTGGCCCCACCCCCGGGAAGGTGCAGCAGAGCTGTGGCTCCCCACACAGCCCGGCCA
GCACCTGTGCTCTGGGCATGGCTGTGCTCCTGGAACGTTCCCTGTCCTGGCTGGTCAGGGGGTGCCCCTGCCAAGAATCG
ACAACTTTATCACAGAGGGAAGGGCCAATCTGTGGAGGCCACAGGGCCAGCTTCTGCCTGGAGTCAGGGCAGGTGGTGGC

ACAAGCCTCGGGGCTGTACCAAAGGGCAGTCGGGCACCACAGGCCCGGGCCTCCACCTCAACAGGCCTCCCGAGCCACTG
GGAGCTGAATGCCAGGAGGCCGAAGCCCTCGCCCCATGAGGGCTGAGAAGGAGTGTGAGCATTTGTGTTACCCAGGGCCG
AGGCTGCGCGAATTACCGTGCACACTTGATGTGAAATGAGGTCGTCGTCTATCGTGGAAACCCAGCAAGGGCTCACGGGA
GAGTTTTCCATTACAAGGTCGTACCATGAAAATGGTTTTTAACCCGAGTGCTTGCGCCTTCATGCTCTGGCAGGGAGGGC
AGAGCCACAGCTGCATGTTACCGCCTTTGCACCAGCTCCAGAGGCTTGGGACCAGGCTGTCTCAGTTCCAGGGTGCGTCC
GGCTCAGACCGCCCTCCTCTCTGCCTTCTCTCTCTGCCTCAAATCTTCCCTCGTTTGCATCTCCCTGACGCGTGCCTGGG
CCCTCGTGCAAGCTGCTTGACTCCTTTCCGGAAACCCTTGGGGTGTGCTGGATACAGGTGCCACTGAGGACTGGAGGTGT
CTGACACTGTGGTTGACCCCAGGGTCCAGCTGGCGTGCTTGGGGCCTCCTTGGGCCATGATGAGGTCAGAGGAGTTTTCC
CAGGTGAAAACTCCTGGGAAACTCCCAGGGCCATGTGACCTGCCACCTGCTCCTCCCATATTCAGCTCAGTCTTGTCCTC
ATTTCCCCACCAGGGTCTCTAGCTCCGAGGAGCTCCCGTAGAGGGCCTGGGCTCAGGGCAGGGCGGCTGAGTTTCCCCAC
CCATGTGGGGACCCTTGGGTAGTCGCTTGATTGGGTAGCCCTGAGGAGGCCGAGATGCGATGGGCCACGGGCCGTTTCCA
AACACAGAGTCAGGCACGTGGAAGGCCCAGGAATCCCCTTCCCTCGAGGCAGGAGTGGGAGAACGGAGAGCTGGGCCCCG
ATTTCACGGCAGCCAGGCTGCAGTGGGCGAGGCTGTGGTGGTCCACGTGGCGCTGGGGGCGGGGTCTGATTCAAATCCGC
TGGGGCTCGGCCTTCCTGGCCCGTGCTGGCCGCGCCTCCACACGGGCTTGGGGTGGACGCCCCGACCTCTAGCAGGTGGC
TATTTCTCCCTTTGGAAGAGAGCCCCTCACCCATGCTAGGTGTTTCCCTCCTGGGTCAGGAGCGTGGCCGTGTGGCAACC
CCGGGACCTTAGGCTTATTTATTTGTTTAAAAACATTCTGGGCCTGGCTTCCGTTGTTGCTAAATGGGGAAAAGACATCC
CACCTCAGCAGAGTTACTGAGAGGCTGAAACCGGGGTGCTGGCTTGACTGGTGTGATCTCAGGTCATTCCAGAAGTGGCT
CAGGAAGTCAGTGAGACCAGGTACATGGGGGGCTCAGGCAGTGGGTGAGATGAGGTACACGGGGGGCTCAGGCAGTGGGT
GAGGCCAGGTACATGGGGGGCTCAGGCACTGGGTGAGATGAGGTACACGGGGGGCTCAGGCAGAGGGTCAGACCAGGTAC
ACGGGGGCTCTGATCACACGCACATATGAGCACATGTGCACATGTGCTGTTTCATGGTAGCCAGGTCTGTGCACACCTGC
CCCAAAGTCCCAGGAAGCTGAGAGGCCAAAGATGGAGGCTGACAGGGCTGGCGCGGTGGCTCACACCTGTAGTCCCAGCA
CTTTGGGAGGCCGAGGCGAGAGGATCCCTTGAGCCCAGGAGTTTAAGACCAGCCTGAGCAACATAGTAGAACCCCATCTC
TATGAAAAATAAAAACAAAAATTAGCTGAACATGGTGGTGTGCGCCTGTAGTTCCAATACTTGGGAGGCTGAAGTGGGAG
GATCACTTGAGCCCAGGAGGTGGAAGCTGCAGTGAGCTGAGATTGCACCACTGTACTGCAGCCTGGGTGACAGAGTGAGA
GCCCATCTCAACAACAACAAAGAAGACTGACAAATGCAGTTTCTTGGAAAGAAACATTTAGTAGGAACTTAACCTACACA
CAGAAGCCAAGTCGGTGTCTCGGTGTCAGTGAGATGAGATGATGGGTCCTCACACCATCACCCCAGACCCAGGGTTTATG
CACCACAGGGGCGGGTGGCTCAGAAGGGATGCGCAGGACGTTGATATACGATGACATCAAGGTTGTCTGACGAAGGGCAG
GATTCATGATAAGTACCTGCTGGTACACAAGGAACAATGGATAAACTGGAAACCTTAGAGGCCTTCCCGGAACAGGGGCT
AATCAGAAGCCAGCATGGGGGGCTGGCATCCAGGATGGAGCTGCTTCAGCCTCCACATGCGTGTTCATACAGATGGTGCA
CAGAAACGCAGTGTACCTGTGCACACACAGACACGCAGCTACTCGCACACACAAGCACACACACAGACATGCATGCATGC
ATCCGTGTGTGTGCACCTGTGCCCATGAGGAAACCCATGCATGTGCATTCATGCACGCACACAGGCACCGGTGGGCCCAT
GCCCACACCCACGAGCACCGTCTGATTAGGAGGCCTTTCCTCTGACGCTGTCCGCCATCCTCTCAG

Intron 14 (SEQ ID NO 18)

GTATGTGCAGGTGCCTGGCCTCAGTGGCAGCAGTGCCTGCCTGCTGGTGTTAGTGTGTCAGGAGACTGAGTGAATCTGGG
CTTAGGAAGTTCTTACCCCTTTTCGCATCAGGAAGTGGTTTAACCCAACCACTGTCAGGCTCGTCTGCCCGCCCTCTCGT
GGGGTGAGCAGAGCACCTGATGGAAGGGACAGGAGCTGTCTGGGAGCTGCCATCCTTCCCACCTTGCTCTGCCTGGGGAA
GCGCTGGGGGGCCTGGTCTCTCCTGTTTGCCCCATGGTGGGATTTGGGGGGCCTGGCCTCTCCTGTTTGCCCTGTGGTGG
GATTGGGCTGTCTCCCGTCCATGGCACTTAGGGCCCTTGTGCAAACCCAGGCCAAGGGCTTAGGAGGAGGCCAGGCCCAG
GCTACCCCACCCCTCTCAGGAGCAGAGGCCGCGTATCACCACGACAGAGCCCCGCGCCGTCCTCTGCTTCCCAGTCACCG

TCCTCTGCCCCTGGACACTTTGTCCAGCATCAGGGAGGTTTCTGATCCGTCTGAAATTCAAGCCATGTCGAACCTGCGGT
CCTGAGCTTAACAGCTTCTACTTTCTGTTCTTTCTGTGTTGTGGAAATTTCACCTGGAGAAGCCGAAGAAAACATTTCTG
TCGTGACTCCTGCGGTGCTTGGGTCGGGACAGCCAGAGATGGAGCCACCCCGCAGACCGTCGGGTGTGGGCAGCTTTCCG
GTGTCTCCTGGGAGGGGAGCTGGGCTGGGCCTGTGACTCCTCAGCCTCTGTTTTCCCCCAG

Intron 15 (WEQ ID NO 19)

GCAAGTGTGGGTGGAGGCCAGTGCGGGCCCCACCTGCCCAGGGGTCATCCTTGAACGCCCTGTGTGGGGCGAGCAGCCTC
AGATGCTGCTGAAGTGCAGACGCCCCCGGGCCTGACCCTGGGGGCCTGGAGCCACGCTGGCAGCCCTATGTGATTAAACG
CTGGTGTCCCCAGGCCACGGAGCCTGGCAGGGTCCCCAACTTCTTGAACCCCTGCTTCCCATCTCAGGGGCGATGGCTCC
CCACGCTTGGGAGCCTTCTGACCCCTGACCTGTGTCCTCTCACAGCCTCTTCCCTGGCTGCTGCCCTGAGCTCCTGGGGT
CCTGAGCAAGTTCTCTCCCCGCCCCGCCGCTCCAGCGTCACTGGGCTGCCTGTCTGCTCGCCCCGGTGGAGGGGTGTCTG
TCCCTTCACTGAGGTTCCCACCAGCCAGGGCCACGAGGTGCAGGCCCTGCCTGCCCGGCCACCCACACGTCCTAGGAGGG
TTGGAGGATGCCACCTCTGGCCTCTTCTGGAACGGAGTCTGATTTTGGCCCCGCAG

3'-untranskribierter Bereich (SEQ ID NO 20)

ATCTCATGTTTGAATCCTAATGTGCACTGCATAGACACCACTGTATGCAATTACAGAAGCCTGTGAGTGAACGGGGTGGT
GGTCAGTGCGGGCCCATGGCCTGGCTGTGCATTTACGGAAGTCTATGAGTGAATGGGGTTGTGGTCAGTGCGGGCCCATG
GCCTGGCTGGGCCTGGGAGGTTTCTGATGCTGTGAGGCAGGAGGGGAAGGAGGGTAGGGGATAGACAGTGGGAGCCCCCA
CCCTGGAAGACATAACAGTAAGTCCAGGCCCGAAGGGCAGCAGGGATGCTGGGGGCCCAGCTTGGGCGGCGGGGATGATG
GAGGGCCTGGCCAGGGTGGCAGGGATGATGGGGGCCCCAGCTGGGGTGGCAGGGGTGATGGGGGGGGCTGGTCTGGGTGG
CGGGGAAGATGGGGAAGCCTGGCTGGGCCCCCTCCTCCCCTGCCTCCCACCTGCAGCCGTGGATCCGGATGTGCTTCCCT
GGTGCACATCCTCTGGGCCATCAGCTTTCATGGAGGTGGGGGGCAGGGGCATGACACCATCCTGTATAAAATCCAGGATT
CCTCCTCCTGAACGCCCCAACTCAGGTTGAAAGTCACATTCCGCCTCTGGCCATTCTCTTAAGAGTAGACCAGGATTCTG
ATCTCTGAAGGGTGGGTAGGGTGGGGCAGTGGAGGGTGTGGACACAGGAGGCTTCAGGGTGGGGCTGGTGATGCTCTCTC
ATCCTCTTATCATCTCCCAGTCTCATCTCTCATCCTCTTATCATCTCCCAGTCTCATCTGTCTTCCTCTTATCTCCCAGT
CTCATCTGTCATCCTCTTACCATCTCCCAGTCTCATCTCTTATCCTCTTATCTCCTAGTCTCATCCAGACTTACCTCCCA
GGGCGGGTGCCAGGCTCGCAGTGGAGCTGGACATACGTCCTTCCTCAGGCAGAAGGAACTGGAAGGATTGCAGAGAACAG
GAGGGGCGGCTCAGAGGGACGCAGTCTTGGGGTGAAGAAACAGCCCCTCCTCAGAAGTTGGCTTGGGCCACACGAAACCG
AGGGCCCTGCGTGAGTGGCTCCAGAGCCTTCCAGCAGGTCCCTGGTGGGGCCTTATGGTATGGCCGGGTCCTACTGAGTG
CACCTTGGACAGGGCTTCTGGTTTGAGTGCAGCCCGGACGTGCCTGGTGTCGGGGTGGGGGCTTATGGCCACTGGATATG
GCGTCATTTATTGCTGCTGCTTCAGAGAATGTCTGAGTGACCGAGCCTAATGTGTATGGTGGGCCCAAGTCCACAGACTG
TGTCGTAAATGCACTCTGGTGCCTGGAGCCCCCGTATAGGAGCTGTGAGGAAGGAGGGGCTCTTGGCAGCCGGCCTGGGG
GCGCCTTTGCCCTGCAAACTGGAAGGGAGCGGCCCCGGGCGCCGTGGGCGGACGACCTCAAGTGAGAGGTTGGACAGAAC
AGGGCGGGGACTTCCCAGGAGCAGAGGCCGCTGCTCAGGCACACCTGGGTTTGAATCACAGACCAACaGGTCAGGCCATT
GTTCAGCTATCCATCTTCTACAAAGCTCCAGATTCCTGTTTCTCCGGGTGTTTTTTGTTGAAATTTTACTCAGGATTACT
TATATTTTTTGCTAAAGTATTAGACCCTTAAAAAAGGTATTTGCTTTGATATGGCTTAACTCACTAAGCACCTACTTTAT
TTGTCTGTTTTTATTTATTATTATTATTATTATTAGAGATGGTGTCTACTCTGTCACCCAGGTTGTTAGTGCAGTGGCAC
AGTCATGGCTCGCTGTAGCCGCAAACCCCCAGGCTCAAGTGATCCTCCGGCCTCAGCTTCCCAGAGTGCTGGGATTACAG
GTGTGAGCCACTGCCCTTGCCTGGCACTTTTAAAAACCACTATGTAAGGTCAGGTCCAGTGGCTTCCACACCTGTCATCC
CAGTAGTTTGGGAAGCCGAGGCAGAAGGATTGTCTGAGGCCAGGAGTTTGAGACCAGCATGGGTAACATAGGGAGACCCC

ATCTCTACAAAAAATGCAAAAAGTTATCCGGGCGTGGGGTCCAGCATCTGTAGTCCCAGCTGCTCGGGAGGCTGAGTGGG
AGGATCGCTTGAGCCCGGGAGGTCATGGCTGCAGTGAGCTGTGATTGTACCATCGCACTCCAGCCTGGGCAACAGAGTGA
GACCCTGTCTCAAAAAAAAAAAAAAAAAAAAGAAGGAGAAGGAGAAGAGAAGAAGGAAGAAGGAAAGAGAAGAAGAAG
GAAGAAGGAAGAAAGAAGGAGAAGGAGGCCTGCTAGGTGCTAGGTAGACTGTCAAATCTCAGAGCAAAATGAAAATAACA
AAGTTTTAAAGGGAAAGAAAAACCCCAGCTCTTTGGACTTCCTTAGGCCTGAACTTCATCTCAAGCAGCTTCCTTCCACA
GACAAGCGTGTATGGAGCGAGTGAGTTCAAAGCAGAAAGGGAGGAGAAGCAGGCAAGGGTGGAGGCTGTGGGTGACACCA
GCCAGGACCCCTGAAAGGGAGTGGTTGTTTTCCTGCCTCAGCCCCACGCTCCTGCCGGTCCTGCACCTGCTGTAACCGTC
GATGTTGGTGCCAGGTGCCCACCTGGGAAGGATGCTGTGCAGGGGGCTTGCCAAACTTTGGTGGGTTTCAGAAGCCCCAG
GCACTTGTGGCAGGCACAATTACAGCCCCTCCCCAAAGATGCCCACGTCCTTCTCCTGGAACCTGTGAATGTGTCACCCG
CAAGGCAGAGGCTGGTGAAGGCTGCAGGTGGAATCACGGCTGCCAGTCAGCCGATCTTAAGGTCATCCTGGATTATCTGG
TGGGCCTGATATGGCCACAAGGGTCCCTAGAAGTGAGAGAGGGAGGCAGGGGAGAGTCAGAGAGGGGACGTGAGAAGGAC
CACTGGCCACTGCTGGCTTTGAGATGGAGGAGGGGGTCCCCAGCCAAGGAATGGGGGCAGCCGCTCCATGCTGGAAAAGC
AAGCAATCCTCCCCGGTCCTGAGGGCACACGGCCCTGCCCACGCCTCGATTTCAGGCCAGTGGGACCTGTTTCAGCTTTC
CGGCCTCCAGAGCTGTAAGATGATGCGTTTGTGTTCAGCCACTAAGCTGCAGTGATTCGTCACAGCAGCAAATGGAATAG
CAGTACAGGGAAATGAATACAGGGACAGTTCTCAGAGTGACTCTCAGCCCACCCCTGGG

Die Charakterisierung der Exons zeigte interessanterweise, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen, funktionell wichtigen hTC-Protein-Domänen auf separaten Exons angeordnet sind. Das Telomerase-charakteristische T-Motiv befindet sich auf Exon 3. Die für die katalytische Funktion der Telomerase wichtigen RT (Reverse-Transkriptase)-Motive 1-7 liegen auf folgenden Exons: RT Motiv 1 und 2 auf Exon 4, RT Motiv 4 auf Exon 9, RT Motiv 5 auf Exon 10, RT Motiv 6 und 7 auf Exon 11. RT Motiv 3 liegt verteilt auf Exon 5 und 6 vor (s. Fig. 8).

Die Aufklärung der Exon-Intron-Struktur des hTC-Gens zeigt auch, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen vier Deletions- bzw. Insertions-Varianten der hTC-cDNA ebenso wie drei weitere, in der Literatur (Kilian et al., 1997) beschriebene hTC-Insertions-Varianten höchstwahrscheinlich alternative Splice-Produkte darstellen. Wie in Fig. 8 gezeigt, lassen sich die Splice Varianten in zwei Gruppen einteilen: Deletionsvarianten und Insertionsvarianten.

Den hTC-Varianten der Deletionsgruppe fehlen spezifische Sequenzabschnitte. Die 36 bp in frame Deletion in Variante DEL1 resultiert höchstwahrscheinlich aus der Benutzung einer alternativen 3'-Splice Akzeptorsequenz in Exon 6, wodurch ein Teil des RT Motivs 3 verlorengeht. In Variante DEL2 werden die normalen 5'-Splice Donor- und 3'-Splice-Akzeptor Sequenzen von Intron 6, 7 und 8 nicht benutzt. Stattdessen wird Exon 6 direkt an Exon 9 fusioniert, wodurch eine Verschiebung des offenen Leserahmens entsteht und in Exon 10 ein Stopcodon auftritt. Variante Del3 stellt eine Kombination aus Variante 1 und 2 dar.

Die Gruppe der Insertions-Varianten zeichnet sich durch die Insertion von Intronsequenzen aus, die zu vorzeitigen Translationsstop führen. Anstelle der normalerweise benutzten 5'-Splice Donorsequenz von Intron 5 wird eine alternative, 3'-lokalisierte Splicestelle in Variante INS1 benutzt, wodurch eine Insertion der ersten 38 bp aus Intron 4 zwischen Exon 4 und Exon 5 entsteht. Ebenso resultiert die

Insertion eines Intron 11-Sequenzbereichs in Variante INS2 aus der Benutzung einer alternativen 5'-Splice Donorsequenz in Intron 11. Da diese Variante in der Literatur (Kilian et al., 1997) nur unzureichend beschrieben wurde, läßt sich die genaue alternative 5'-Splice Donorsequenz dieser Variante nicht bestimmen. Die Insertion von Intron 14 Sequenzen zwischen Exon 14 und Exon 15 in Variante INS3 entsteht durch die Benutzung von einer alternativen 3'-Splice Akzeptorsequenz, wodurch der 3'-Teil von Intron 14 nicht gesplict wird.

Die in unserer Patentanmeldung PCT/EP/98/03469 beschriebene hTC-Variante INS4 (Variante 4) zeichnet sich durch den Ersatz von Exon 15 und dem 5'-Teilbereich von Exon 16 durch die ersten 600 bp des Introns 14 aus. Diese Variante ist auf den Gebrauch einer alternativer internen 5'-Splice Donorsequenz in Intron 14 und einer alternativen 3'-Splice Akzeptorsequenz in Exon 16 zurückzuführen, woraus ein veränderter C-Terminus resultiert.

Die *in vivo*-Generation wahrscheinlich nicht-funktioneller hTC-Proteinvarianten, die mit der Funktion des vollständigen hTC-Proteins interferieren könnten, stellt zusätlich zur Transkriptionsregulation einen möglichen Mechanismus dar, um die hTC-Proteinfunktion zu kontrollieren. Bis heute ist die Funktion der hTC-Splicevarianten nicht bekannt. Obwohl die meisten dieser Varianten vermutlich für Proteine ohne Reverse-Transkriptase-Aktivität kodieren, könnten sie dennoch eine entscheidende Rolle als transdominant-negative Telomerase-Regulatoren spielen, indem sie z.B. um die Interaktion mit wichtigen Bindungspartnern kompetieren.

Die Suche nach möglichen Transkriptionsfaktorbindungstellen wurde mit dem „Find Pattern"-Algorithmuses aus dem „GCG Sequenz Analysis" Programmpacket der „Genetics Computer Group" (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen für Transkriptionsfaktoren in der Nukleotidsequenz von Intron 2 identifiziert, die in der Tab. 2 aufgelistet sind. Darüberhinaus wurde im Intron 1 eine Sp1-Bindungsstelle (Pos. 43) und im 5'-

untranslatiertem Bereich eine c-Myc-Bindungsstelle (cDNA-Position 29-34, vergl. Fig. 6) gefunden.

Beispiel 6

Um den oder die Startpunkt(e) der hTC-Transkription in HL 60 Zellen zu ermitteln, wurde das 5'-Ende der hTC-mRNA durch Primer-Extension-Analyse bestimmt.

Es wurden 2 µg PolyA$^+$-RNA aus HL-60-Zellen für 10 min bei 65°C denaturiert. Zur Primeranlagerung wurden 1 µl RNasin (30-40 U/ml) und 0,3-1 pmol radioaktiv markierter Primer (5'GTTAAGTTGTAGCTTACACTGGTTCTC 3'; 2,5-8x10^5 cpm) zugegeben und für 30 min bei 37°C in einem Gesamtvolumen von 20 µl inkubiert. Nach Zugabe von 10 µl 5xReverse Transkriptase-Puffer (Fa. Gibco-BRL), 2 µl 10 mM dNTPs, 2 µl RNasin (s.o.), 5µl 0,1 M DTT (Fa. Gibco-BRL) 2 µl ThermoScript RT (15 U/µl; Fa. Gibco-BRL) und 9 µl DEPC-behandeltes Wasser erfolgte die Primer-Verlängerung in einem Gesamtvolumen für 1 h bei 58°C. Die Reaktion wurde durch 4 µl 0,5 M EDTA, pH 8,0, gestoppt und die RNA nach Zugabe von 1 µl RNaseA (10 mg/ml) für 30 min bei 37°C abgebaut. Hierauf wurden 2,5 µg gescherte Kalbsthymus-DNA und 100 µl TE addiert und einmal mit 150µl Phenol/Cloroform (1:1) extrahiert. Die DNA wurde unter Zusatz von 15 µl 3 M Na-Acetat und 450 µl Ethanol für 45 min bei –70°C gefällt und anschließend für 15 min bei 14000 Upm abzentrifugiert. Das Präzipitat wurde einmal mit 70 %igem Ethanol gewaschen, luftgetrocknet und in 8 µl Sequenzierungs-Stoplösung gelöst. Nach 5 min Denaturierung bei 80°C wurden die Proben auf ein 6 %iges Polyacrylamidgel aufgetragen und elektrophoretisch (Ausubel et al., 1987) aufgetrennt (Fig. 5).

Hierbei wurde eine Haupt-Transkriptionsstartstelle identifiziert, die 1767 bp 5' vom ATG-Startcodon der hTC-cDNA Sequenz lokalisiert ist (Nukleotidposition 3346 in Fig. 4). Die Nukleotidsequenz um diesen Haupttranskriptionsstart ($TTA_{+1}TTGT$) repräsentiert darüberhinaus ein Initiator-Element (Inr), das in 6 von 7 Nukleotiden

mit dem Konsensusmotiv ($PyPyA_{+1}Na/tPyPy$) (Smale, 1997) eines Initiator-Elementes übereinstimmt.

In unmittelbarer Nähe des experimentell identifizierten Haupt-Transkriptionsstartes konnte keine eindeutige TATA-Box identifiziert werden, so daß der hTC-Promoter wahrscheinlich in die Familie der TATA-losen Promotoren (Smale, 1997) einzuordnen ist. Allerdings wurde durch Bioinformatik Analyse eine potentielle TATA-Box von Nukleotidposition 1306 bis 1311 (Fig. 4) gefunden. Die zusätzlich um den Haupt-Transkriptionsstart beobachteten Neben-Transkriptionsstarts wurden auch bei anderen TATA-losen Promotoren beschrieben (Geng and Johnson, 1993), wie z.B. in den stark regulierten Promotoren einiger Zellzyklusgene (Wick *et al.*, 1995).

Beispiel 7

Zusätzlich zu dem in Beispiel 6 beschriebenen, in HL60 Zellen identifizierten Startpunkt des hTC Transkriptes, wurde ein weiterer Transkriptionsstartbereich in HL60 Zellen identifiziert. Anhand von RT-PCR-Analysen wurde die Region des Transkriptionsstarts des hTC-Gens in HL60 Zellen auf die bp –60 bis –105 eingegrenzt.

Unter Einsatz von 0,4 µg Poly A-RNA aus HL60 Zellen (Clontech) und dem genspezifischen Primer GSP13 (5'-CCTCCAAAGAGGTGGCTTCTTCGGC-3', cDNA-Position 920-897) wurde hierfür die cDNA mit Hilfe des „First Strand cDNA-Synthesis Kit" (Clontech) nach Angaben der Hersteller synthetisiert. In einem Endvolumen von 50 µl wurden 1 µl cDNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer F (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 10 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 36 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer

angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Insgesamt wurden sechs verschiedene 5'-PCR Primer (Primer HTRT5B: 5'-CGCAGCCACTACCGCGAGGTGC-3', cDNA-Position 105 bis 126; Primer C5S: 5'-CTGCGTCCTGCTGCGCACGTGGGAAGC-3', 5'-flankierende Region -49 bis -23; Primer PRO-TEST1: 5'-CTCGCGGCGCGAGTTTCAGGCAG-3', 5'-flankierende Region –74 bis -52; Primer PRO-TEST2: 5'-CCAGCCCCTCCCCTTCCTTTCC-3', 5'-flankierende Region –112 bis -91; Primer PRO-TEST4: 5'-CCAGCTCCGCCTCCTCCGCGC-3', 5'-flankierende Region –191 - -171; Primer RP-3A: 5'-CTAGGCCGATTCGACCTCTCTCC-3', 5'-flankierende Region –427 bis -405) mit dem 3'-PCR Primer C5Rrück (5'-GTCCCAGGGCACGCACACCAG-3', cDNA-Position 245 bis 225) kombiniert. Als Kontrolle wurde zusätzlich zu den Oligo-dT- und GSP13-geprimten cDNAs auch genomische DNA für die PCR eingesetzt. Wie in Fig. 9 gezeigt, wurde nur mit den Primerkombinationen HTRT5B-C5Rrück, C5S-C5Rrück und PRO-TEST1-C5Rrück ein PCR-Produkt erhalten, was darauf hinweist, daß der Startpunkt der hTC-Transkription in der Region zwischen bp-60 und bp-105 liegt.

Beispiel 8

In der ca. 11,2 kb isolierten 5'-flankierenden Region des hTC-Gens befinden sich mehrere extrem GC-reiche Bereiche, sog. CpG Islands. Ein CpG Islands mit einem GC-Gehalt von > 70 % reicht von bp – 1214 bis in Intron 2. Zwei weitere GC-reiche Bereiche mit einem GC-Gehalt von > 60 % reichen von bp –3872 bis bp –3113 bzw. bp –5363 bis bp –3941. Die Lage der CpG Islands ist in der Fig. 11 graphisch dargestellt.

Die Suche nach möglichen Transkriptionsfaktorbindungstellen wurde mit dem „Find Pattern"-Algorithmuses aus dem „GCG Sequenz Analysis" Programmpacket der „Genetics Computer Group" (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen in der Region bis -900 bp upstream vom

Translations-Startcodon ATG indentifiziert: fünf Sp1-Bindungsstellen, eine c-Myc-Bindungsstelle, eine CCAC-Box (Fig. 10). Zusätzlich wurden eine CCAAT-Box und eine zweite c-Myc-Bindungsstelle an den Positionen –1788 bzw. –3995 der 5'-flankierenden Region gefunden.

Beispiel 9

Um die Aktivität des hTC-Promotors zu analysieren, wurden durch PCR-Amplifikation vier verschieden lange hTC-Promotorsequenzabschnitte generiert und 5' vor das Reportergen Luziferase in den Vektor pGL2 der Fa. Promega kloniert. Als DNA-Quelle für die PCR-Amplifikation wurde das aus dem Phagenklon P12 subklonierte, 8,5 kb große SacI-Fragment gewählt. In einem Endvolumen von 50 µl wurden 35 ng dieser DNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 20 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 30 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Als 3'-PCR-Primer wurde jeweils der Primer PK-3A (5'-GCAAGCTTGACGCAGCGCTGCCTGAAACTCG-3', Position –43 bis -65) gewählt, der einen Sequenzbereich 42 bp upstream vom START-Codon ATG erkennt. Durch Kombination des PK-3A-Primers mit dem 5'-PCR-Primer PK-5B (5'-CCAGATCTCTGGAACACAGAGTGGCAGTTTCC-3', Position –4093 bis -4070) wurde ein 4051 bp großes Promotor-Fragment amplifiziert (NPK8). Die Kombination des Primerpaares PK-3A und PK-5C (5'-CCAGATCTGCATGAAGTGTGTGGGGATTTGCAG-3', Position –3120 bis-3096) führte zur Amplifikation eines 3078 bp großen Promotorfragmentes (NPK15). Ein 2068 bp großes Promotorfragment wurde durch die Verwendung der Primerkombination PK-3A und PK-5D (5'-

GGAGATCTGATCTTGGCTTACTGCAGCCTCTG-3', Position –2110 bis -2087) amplifziert (NPK22). Der Einsatz der Primerkombination PK-3A und PK-5E (5'-GGAGATCTGTCTGGATTCCTGGGAAGTCCTCA-3', Position –1125 bis -1102) führte schließlich zur Amplifikation eines 1083 bp großen Promotorfragmentes (NPK27). Der PK-3A Primer enthält eine HindIII Erkennungssequenz. Die verschiedenen 5'-Primer enthalten eine BglII-Erkennungssequenz.

Die entstandenen PCR-Produkte wurden mit Hilfe des QIA quick spin PCR Purification Kits der Fa. Qiagen nach Angaben der Hersteller aufgereinigt und anschließend mit den Restriktionsenzymen BglII und HindIII verdaut. Mit den gleichen Restriktionsenzymen wurde der pGL2-Promotor-Vektor verdaut und der in diesem Vektor enthaltene SV40-Promotor freigesetzt und abgetrennt. Die PCR-Promotorfragmente wurden in den Vektor ligiert, in kompetente DH5α-Bakterien der Fa. Gibco/BRL transformiert. Aus transformierten Bakterienklonen wurde DNA für die nachfolgend beschriebenen Promotor-Aktivitäts-Analysen mit Hilfe des Qiagen Plasmid-Kits der Fa. Qiagen isoliert.

Beispiel 10

Die Aktivität des hTC-Promotors wurde in transienten Transfektionen in eukaryotischen Zellen analysiert.

Alle Arbeiten mit eukaryotischen Zellen erfolgten an einem sterilen Arbeitsplatz. CHO-K1 und HEK 293 Zellen wurden von der American Type Culture collection bezogen.

CHO-K1 Zellen wurden in DMEM Nut Mix F-12 Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 21331-020) mit 0,15 % Streptomycin/Penezillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) gehalten.

HEK 293 Zellen wurden in DMOD Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 41965-039) mit 0,15 % Streptomycin/Penizillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) kultiviert.

CHO-K1 und HEK 293 Zellen wurden in wasssergesättigter Atmosphäre bei 37°C unter Begasung mit 5 % CO_2 kultiviert. Bei konfluentem Zellrasen wurde das Medium abgesaugt, die Zellen mit PBS (100 mM KH_2PO_4 pH 7,2; 150 mM NaCl) gewaschen und durch Zugabe einer Trypsin-EDTA Lösung (Fa. Gibco-BRL) abgelöst. Das Trypsin wurde durch Mediumzugabe inaktiviert und die Zellzahl mit einer Neubauer-Zählkammer ermittelt, um die Zellen in gewünschter Dichte auszuplattieren.

Für die Transfektion wurden pro Well jeweils 2x 10^5 –HEK 293 Zellen in einer 24-well Zellkulturplatte ausplattiert. Nach 3 Stunden wurde das HEK 293 Medium entfernt. Für die Transfektion wurden bis zu 2,5 µg Plasmid-DNA, 1 µg eines CMV ß-Gal Plasmidkonstruktes (Fa. Stratagene, Bestellnummer: 200388), 200 µl serumfreies Medium und 10 µl Transfektionsreagenz (DOTAP der Fa. Boehringer Mannheim) für 15 Minuten bei Raumtemperatur inkubiert und anschließend auf die HEK 293 Zellen gleichmäßig aufgetropft. Nach 3 Stunden wurden 1,5 ml Medium hinzugegeben. Nach 20 Stunden wurde das Medium gewechselt. Nach weiteren 24 Stunden wurden die Zellen zur Bestimmung der Luziferase- und der ß-Gal-Aktivität geerntet. Dazu wurden die Zellen im Zellkultur-Lysisreagenz (25 mM Tris [pH 7,8] mit H_3PO_4; 2 mM CDTA; 2 mM DTT; 10% Glycerol; 1% Triton X-100) für 15 Minuten bei Raumtemperatur lysiert. Zwanzig µl dieses Zellysats wurden mit 100 µl Luziferase-Assaypuffer (20 mM Tricin; 1,07 mM $(MgCO_3)_4$ $Mg(OH)_2{\cdot}5H_2O$; 2,67 mM $MgSO_4$; 0,1 mM EDTA; 33,3 mM DTT; 270 µM Coenzym A; 470 µM Luciferin, 530 µM ATP) gemischt und das durch die Luziferase generierte Licht gemessen.

Zur Messung der ß-Galaktosidaseaktivität wurden gleiche Mengen Zellysat und ß-Galaktosidase-Assaypuffer (100 mM Natriumphosphatpuffer pH 7,3; 1 mM $MgCl_2$;

50 mM ß-Merkaptoethanol; 0,665 mg/ml ONPG) für mindestens 30 Minuten bei 37°C oder bis eine leichte Gelbfärbung auftrat, inkubiert. Die Reaktion wurde durch Zugabe von 100 µl 1 M Na_2CO_3 gestoppt und die Absorption bei 420 nm bestimmt.

Für die Analyse des hTC-Promotors wurden vier verschieden lange hTC-Promotorsequenzabschnitte 5' vor das Reportergen Luziferase kloniert (vergl. Beispiel 9).

In der Fig. 11 sind die relativen Luziferase Aktivitäten zweier unabhängiger Transfektionen mit den Konstrukten NPK8, NPK15, NPK22 und NPK27 in HEK 293 Zellen aufgetragen. Jedes Experiment wurde in Duplikaten durchgeführt. Darüberhinaus wurde die Standardabweichung angegeben. Das Konstukt NPK 27 zeigt eine 40fach höhere Luziferaseaktivität als die Basalaktivität des promotorlosen Luziferase-Kontrollkonstrutes (pGL2-basic) und eine 2 bis 3fach höhere Aktivität als das SV40 Promotorkontroll-Konstrukt (pGL2PRO). Interessanterweise wurde im Vergleich zu dem Konstrukt NPK27 eine 2 bis 3fach geringere Luziferaseaktivität in mit längeren hTC Promotorkonstukten (NPK8, NPK15, NPK22) transfizierten Zellen beobachtet. Ähnliche Ergebnisse wurden auch in CHO Zellen beobachtet (Daten nicht gezeigt).

Literaturvereichnis

Allsopp, R. C., Vazire, H., Pattersson, C., Goldstein, S., Younglai, E.V., Futcher, A.B., Greider, C.W. und Harley, C.B. (1992). Telomere length predicts replicative capacity of human fibroblasts. Proc. Natl. Acad. Sci. 89, 10114-10118.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (1987). Current protocols in molecular biology. Greene Publishing Associates and Whiley-Intersciences, New York.

Blasco, M. A., Rizen, M., Greider, C. W. und Hanahan, D. (1996). Differential regulation of telomerase activity and telomerase RNA during multistage tumorigenesis. Nature Genetics 12, 200-204.

Broccoli, D., Young, J. W. und deLange, T. (1995). Telomerase activity in normal and malignant hematopoietic cells. Proc. Natl. Acad. Sci. 92, 9082-9086.

Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G. Greider, C.W. Harley, C. B. und Bacchetti S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. EMBO J. 11, 1921-1929.

Feng, J., Funk, W. D., Wang, S.-S., Weinrich, S. L., Avilion, A.A., Chiu, C.-P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J., Le, S., West, M.D., Harley, C.B., Andrews, W.H., Greider, C.W. und Villeponteau, B. (1995). The RNA component of human telomerase. Science 269, 1236-1241.

Geng, Y., and Johnson, L.F. (1993). Lack of an initiator element is responsible for multiple transcriptional initiation sites of the TATA less mouse thymidine synthasse promoter. Mol. Cell. Biol 14:4894.

Goldstein, S. (1990). Replicative senescence: The human fibroblast comes of age. Science 249, 1129-1133.

Harley, C.B., Futcher, A.B., Greider, C.W., 1990. Telomeres shorten during ageing of human fibroblasts. Nature 345, 458-460.

Hastie, N. D., Dempster, M., Dunlop, M. G., Thompson, A. M., Green, D.K. und Allshire, R.C. (1990). Telomere reduction in human colorectal carcinoma and with ageing. Nature 346, 866-868.

Hiyama, K., Hirai, Y., Kyoizumi, S., Akiyama, M., Hiyama, E., Piatyszek, M.A., Shay, J.W., Ishioka, S. und Yamakido, M. (1995). Activation of telomerase in human lymphocytes and hematopoietic progenitor cells. J. Immunol. 155, 3711-3715.

Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C. B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L. und Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. Science 266, 2011-2015.

Latchman, D.S. (1991). Eukaryotic transcription factors. Academic Press Limited, London.

Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V. und Cech T.R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. Science 276: 561-567.

Lundblad, V. und Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. Cell 57, 633-643.

McClintock, B. (1941). The stability of broken ends of chromosomes in *Zea mays*. Genetics 26, 234-282.

Meyne, J., Ratliff, R. L. und Moyzis, R. K. (1989). Conservation of the human telomere sequence $(TTAGGG)_n$ among vertebrates. Proc. Natl. Acad. Sci. 86, 7049-7053.

Olovnikov, A. M. (1973). A theory of marginotomy. J. Theor. Biol. 41, 181-190.

Sandell, L. L. und Zakian, V. A. (1993). Loss of a yeast telomere: Arrest, recovery and chromosome loss. Cell 75, 729-739.

Shapiro, M.B., Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucl. Acids Res. 15, 7155-7174.

Smale, S.T. and Baltimore, D. (1989). The „initiator" as a transcription control element. Cell 57:103-113.

Smale, S.T. (1997). Transcription initation from TATA-less promoters within eukaryotic protein-coding genes. Biochimica et Biophysica Acta 1351, 73-88.

Shay, J. W. (1997). Telomerae and Cancer. Ciba Foundation Meeting: Telomeres and Telomerase. London.

Vaziri, H., Dragowska, W., Allsopp, R. C., Thomas, T. E., Harley, C.B. und Landsdorp, P.M. (1994). Evidence for a mitotic clock in human hematopoietic stem cells: Loss of telomeric DNA with age. Proc. Natl. Acad. Sci. 91, 9857-9860.

Wick, M., Härönen, R., Mumberg, D., Bürger, C., Olsen, B.R., Budarf, M.L., Apte, S. S. and Müller, R. (1995). Structure of the human TIMP-3 gene and its cell-cycle-regulated promoter. Biochemical Jornal 311, 549-554.

Zakian, V. A. (1995). Telomeres: Beginning to understand the end. Science 270, 1601-1607.

Patentansprüche

1. Regulatorische DNA-Sequenzen für das Gen der humanen katalytischen Telomerase-Untereinheit.

2. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um Intronsequenzen gemäß SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 und/oder 20 oder um regulatorisch wirksame Fragmente dieser Sequenzen handelt.

3. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit gemäß Fig. 10 (SEQ ID NO 3) oder um regulatorisch wirksame Fragmente dieser DNA-Sequenz handelt.

4. Rekombinantes Konstrukt, enthaltend eine DNA-Sequenz gemäß einem der Ansprüche 1 bis 3.

5. Rekombinantes Konstrukt gemäß Anspruch 4, dadurch gekennzeichnet, daß es weiterhin eine oder mehrere DNA-Sequenzen enthält, die für Polypeptide oder Proteine kodieren.

6. Vektor, enthaltend ein rekombinantes Konstrukt gemäß Anspruch 4 oder 5.

7. Verwendung von rekombinanten Konstrukten bzw. Vektoren gemäß einem der Ansprüche 4 bis 6 zur Herstellung von Arzneimitteln.

8. Rekombinante Wirtszellen, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß einem der Ansprüche 4 bis 6.

9. Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der humanen katalytischen Telomerase-Untereinheit beeinflussen, das folgende Schritte umfaßt:

A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend DNA-Sequenzen gemäß einem der Ansprüche 1 bis 3, funktionell verknüpft mit einem Reportergen,

B. Messung des Substanzeffektes auf die Reportergenexpression.

10. Verfahren zur Identifizierung von Faktoren, die spezifisch an die DNA gemäß einem der Ansprüche 1 bis 3 oder an Fragmente davon binden, dadurch gekennzeichnet, daß man eine Expressions-cDNA-Bibliothek mit einer DNA-Sequenz gemäß einem der Ansprüche 1 bis 3 oder Teilfragmenten unterschiedlichster Länge als Sonde screent.

11. Transgene Tiere, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß Ansprüchen 4 bis 6.

12. Verfahren zur Detektion Telomerase-assoziierter Zustände bei einem Patienten, das folgende Schritte umfaßt:

A. Inkubation eines rekombinanten Konstruktes bzw. Vektors gemäß Ansprüchen 4 bis 6 das bzw. der zusätzlich ein Reportergen enthält mit Körperflüssigkeiten oder zellulären Proben,

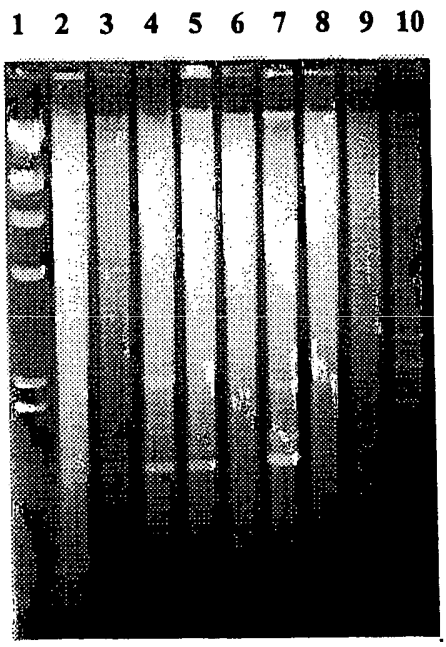
B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten,

C. Vergleich des diagnostischen Wertes mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe.

5

Fig. 1

A



B

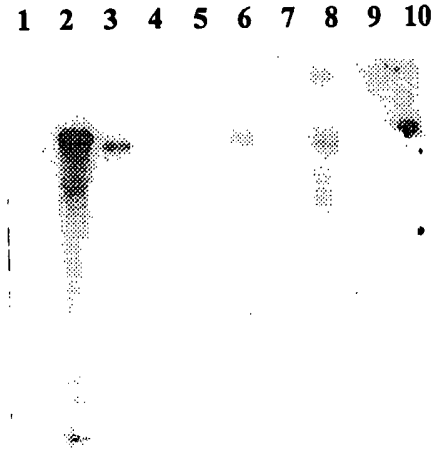


Fig. 2

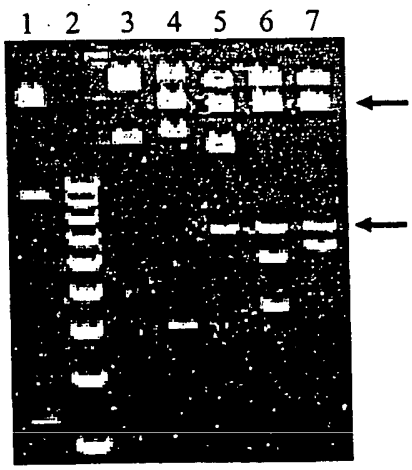


Fig. 3

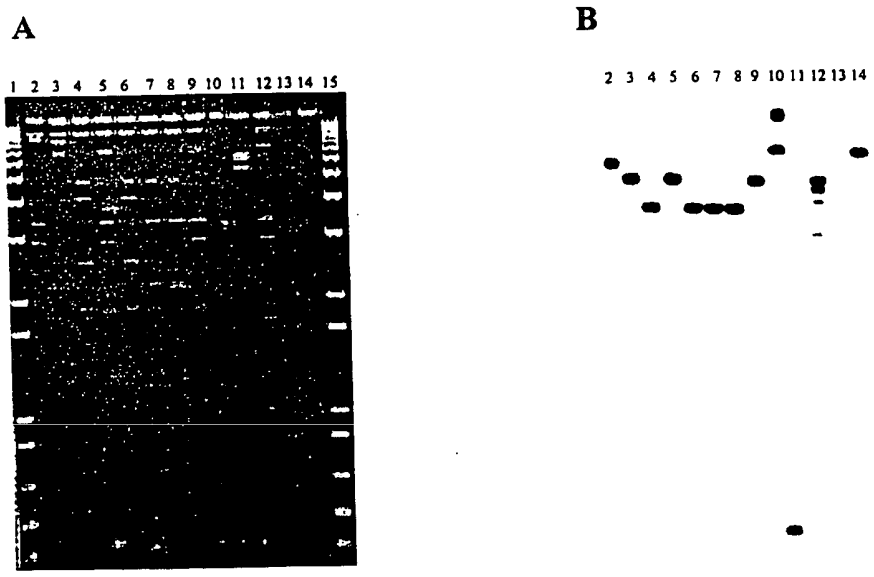


Fig. 4

```
GAGCTCTGAA CCGTGGAAAC GAACATGACC CTTGCCTGCC TGCTTCCCTG GGTGGGTCAA GGGTAATGAA   70
GTGGTGTGCA GGAAATGGCC ATGTAAATTA CACGACTCTG CTGATGGGGA CCGTTCCTTC CATCATTATT  140
CATCTTCACC CCCAAGGACT GAATGATTCC AGCAACTTCT TCGGGTGTGA CAAGCCATGA CAAAACTCAG  210
TACAAACACC ACTCTTTTAC TAGGCCCACA GAGCACGGGC CACACCCCTG ATATATTAAG AGTCCAGGAG  280
AGATGAGGCT GCTTTCAGCC ACCAGGCTGG GGTGACAACA GCGGCTGAAC AGTCTGTTCC TCTAGACTAG  350
TAGACCCTGG CAGGCACTCC CCCAAATTCT AGGGCCTGGT TGCTGCTTCC CGAGGGCGCC ATCTGCCCTG  420
GAGACTCAGC CTGGGGTGCC ACACTGAGGC CAGCCCTGTC TCCACACCCT CCGCCTCCAG GCCTCAGCTT  490
CTCCAGCAGC TTCCTAAACC CTGGGTGGGC CGTGTTCCAG CGCTACTGTC TCACCTGTCC CACTGTGTCT  560
TGTCTCAGCG ACGTAGCTCG CACGGTTCCT CCTCACATGG GGTGTCTGTC TCCTTCCCCA ACACTCACAT  630
GCGTTGAAGG GAGGAGATTC TGCGCCTCCC AGACTGGCTC CTCTGAGCCT GAACCTGGCT CGTGGCCCCC  700
GATGCAGGTT CCTGGCGTCC GGCTGCACGC TGACCTCCAT TTCCAGGCGC TCCCCGTCTC CTGTCATCTG  770
CCGGGGCCTG CCGGTGTGTT CTTCTGTTTC TGTGCTCCTT TCCACGTCCA GCTGCGTGTG TCTCTGCCCG  840
CTAGGGTCTC GGGGTTTTTA TAGGCATAGG ACGGGGGCGT GGTGGGCCAG GGCGCTCTTG GGAAATGCAA  910
CATTTGGGTG TGAAAGTAGG AGTGCCTGTC CTCACCTAGG TCCACGGGCA CAGGCCTGGG GATGGAGCCC  980
CCGCCAGGGA CCCGCCCTTC TCTGCCCAGC ACTTTCCTGC CCCCCTCCCT CTGGAACACA GAGTGGCAGT 1050
TTCCACAAGC ACTAAGCATC CTCTTCCCAA AAGACCCAGC ATTGGCACCC CTGGACATTT GCCCCACAGC 1120
CCTGGGAATT CACGTGACTA CGCACATCAT GTACACACTC CCGTCCACGA CCGACCCCCG CTGTTTTATT 1190
TTAATAGCTA CAAAGCAGGG AAATCCCTGC TAAAATGTCC TTTAACAAAC TGGTTAAACA AACGGGTCCA 1260
TCCGCACGGT GGACAGTTCC TCACAGTGAA GAGGAACATG CCGTTTATAA AGCCTGCAGG CATCTCAAGG 1330
GAATTACGCT GAGTCAAAAC TGCCACCTCC ATGGGATACG TACGCAACAT GCTCAAAAAG AAAGAATTTC 1400
ACCCCATGGC AGGGGAGTGG TTAGGGGGGT TAAGGACGGT GGGGGCGGCA GCTGGGGGCT ACTGCACGCA 1470
CCTTTTACTA AAGCCAGTTT CCTGGTTCTG ATGGTATTGG CTCAGTTATG GGAGACTAAC CATAGGGGAG 1540
TGGGGATGGG GGAACCCGGA GGCTGTGCCA TCTTTGCCAT GCCCGAGTGT CCTGGGCAGG ATAATGCTCT 1610
AGAGATGCCC ACGTCCTGAT TCCCCCAAAC CTGTGGACAG AACCCGCCCG GCCCCAGGGC CTTTGCAGGT 1680
GTGATCTCCG TGAGGACCCT GAGGTCTGGG ATCCTTCGGG ACTACCTGCA GGCCCGAAAA GTAATCCAGG 1750
GGTTCTGGGA AGAGGCGGGC AGGAGGGTCA GAGGGGGGCA GCCTCAGGAC GATGGAGGCA GTCAGTCTGA 1820
GGCTGAAAAG GGAGGGAGGG CCTCGAGCCC AGGCCTGCAA GCGCCTCCAG AAGCTGGAAA AAGCGGGGAA 1890
GGGACCCTCC ACGGAGCCTG CAGCAGGAAG GCACGGCTGG CCCTTAGCCC ACCAGGGCCC ATCGTGGACC 1960
TCCGGCCTCC GTGCCATAGG AGGGCACTCG CGCTGCCCTT CTAGCATGAA GTGTGTGGGG ATTTGCAGAA 2030
GCAACAGGAA ACCCATGCAC TGTGAATCTA GGATTATTTC AAAACAAAGG TTTACAGAAA CATCCAAGGA 2100
CAGGGCTGAA GTGCCTCCGG GCAAGGGCAG GGCAGGCACG AGTGATTTTA TTTAGCTATT TTATTTTATT 2170
TACTTACTTT CTGAGACAGA GTTATGCTCT TGTTGCCCAG GCTGGAGTGC AGCGGCATGA TCTTGGCTCA 2240
CTGCAACCTC CGTCTCCTGG GTTCAAGCAA TTCTCGTGCC TCAGCCTCCC AAGTAGCTGG GATTTCAGGC 2310
GTGCACCACC ACACCCGGCT AATTTTGTAT TTTTAGTAGA GATGGGCTTT CACCATGTTG GTCAAGCTGA 2380
TCTCAAAATC CTGACCTCAG GTGATCCGCC CACCTCAGCC TCCCAAAGTG CTGGGATTAC AGGCATGAGC 2450
CACTGCACCT GGCCTATTTA ACCATTTTAA AACTTCCCTG GGCTCAAGTC ACACCCACTG GTAAGGAGTT 2520
CATGGAGTTC AATTTCCCCT TTACTCAGGA GTTACCCTCC TTTGATATTT TCTGTAATTC TTCGTAGACT 2590
GGGGATACAC CGTCTCTTGA CATATTCACA GTTTCTGTGA CCACCTGTTA TCCCATGGGA CCCACTGCAG 2660
GGGCAGCTGG GAGGCTGCAG GCTTCAGGTC CCAGTGGGGT TGCCATCTGC CAGTAGAAAC CTGATGTAGA 2730
ATCAGGGCGC AAGTGTGGAC ACTGTCCTGA ATCTCAATGT CTCAGTGTGT GCTGAAACAT GTAGAAATTA 2800
AAGTCCATCC CTCCTACTCT ACTGGGATTG AGCCCCTTCC CTATCCCCCC CCAGGGGCAG AGGAGTTCCT 2870
CTCACTCCTG TGGAGGAAGG AATGATACTT TGTTATTTTT CACTGCTGGT ACTGAATCCA CTGTTTCATT 2940
TGTTGGTTTG TTTGTTTTGT TTTGAGAGGC GGTTTCACTC TTGTTGCTCA GGCTGGAGGG AGTGCAATGG 3010
CGCGATCTTG GCTTACTGCA GCCTCTGCCT CCCAGGTTCA AGTGATTCTC CTGCTTCCGC CTCCCATTTG 3080
GCTGGGATTA CAGGCACCCG CCACCATGCC CAGCTAATTT TTTGTATTTT TAGTAGAGAC GGGGGTGGGT 3150
GGGGTTCACC ATGTTGGCCA GGCTGGTCTC GAACTTCTGA CCTCAGATGA TCCACCTGCC TCTGCCTCCT 3220
AAAGTGCTGG GATTACAGGT GTGAGCCACC ATGCCCAGCT CAGAATTTAC TCTGTTTAGA AACATCTGGG 3290
TCTGAGGTAG GAAGCTCACC CCACTCAAGT GTTGTGGTGT TTTAAGCCAA TGATAGAATT TTTTTATTGT 3360
TGTTAGAACA CTCTTGATGT TTTACACTGT GATGACTAAG ACATCATCAG CTTTTCAAAG ACACACTAAC 3430
TGCACCCATA ATACTGGGGT GTCTTCTGGG TATCAGCAAT CTTCATTGAA TGCCGGGAGG CGTTTCCTCG 3500
CCATGCACAT GGTGTTAATT ACTCCAGCAT AATCTTCTGC TTCCATTTCT TCTCTTCCCT CTTTTAAAAT 3570
TGTGTTTTCT ATGTTGGCTT CTCTGCAGAG AACCAGTGTA AGCTACAACT TAACTTTTGT TGGAACAAAT 3640
TTTCCAAACC GCCCCTTTGC CCTAGTGGCA GAGACAATTC ACAAACACAG CCCTTTAAAA AGGCTTAGGG 3710
ATCACTAAGG GGATTTCTAG AAGAGCGACC TGTAATCCTA AGTATTTACA AGACGAGGCT AACCTCCAGC 3780
GAGCGTGACA GCCCAGGGAG GGTGCGAGGC CTGTTCAAAT GCTAGCTCCA TAAATAAAGC AATTTCCTCC 3850
GGCAGTTTCT GAAAGTAGGA AAGGTTACAT TTAAGGTTGC GTTTGTTAGC ATTTCAGTGT TTGCCGACCT 3920
CAGCTACAGC ATCCCTGCAA GGCCTCGGGA GACCCAGAAG TTTCTCGCCC CCTTAGATCC AAACTTGAGC 3990
AACCCGGAGT CTGGATTCCT GGGAAGTCCT CAGCTGTCCT GCGGTTGTGC CGGGGCCCCA GGTCTGGAGG 4060
GGACCAGTGG CCGTGTGGCT TCTACTGCTG GGCTGGAAGT CGGGCCTCCT AGCTCTGCAG TCCGAGGCTT 4130
GGAGCCAGGT GCCTGGACCC CGAGGCTGCC CTCCACCCTG TGCGGGCGGG ATGTGACCAG ATGTTGGCCT 4200
CATCTGCCAG ACAGAGTGCC GGGGCCCAGG GTCAAGGCCG TTGTGGCTGG TGTGAGGCGC CCGGTGCGCG 4270
GCCAGCAGGA GCGCCTGGCT CCATTTCCCA CCCTTTCTCG ACGGGACCGC CCCGGTGGGT GATTAACAGA 4340
TTTGGGGTGG TTTGCTCATG GTGGGGACCC CTCGCCGCCT GAGAACCTGC AAAGAGAAAT GACGGGCCTG 4410
TGTCAAGGAG CCCAAGTCGC GGGGAAGTGT TGCAGGGAGG CACTCCGGGA GGTCCCGCGT GCCCGTCCAG 4480
GGAGCAATGC GTCCTCGGGT TCGTCCCCAG CCGCGTCTAC GCGCCTCCGT CCTCCCCTTC ACGTCCGGCA 4550
TTCGTGGTGC CCGGAGCCCG ACGCCCCGCG TCCGGACCTG GAGGCAGCCC TGGGTCTCCG GATCAGGCCA 4620
GCGGCCAAAG GGTCGCCGCA CGCACCTGTT CCCAGGGCCT CCACATCATG GCCCCTCCCT CGGGTTACCC 4690
```

Fig. 4 (Fortsetzung)

```
CACAGCCTAG GCCGATTCGA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCCT GGGAGCGCGA 4760
GCGGCGCGCG GGCGGGGAAG CGCGGCCCAG ACCCCCGGGT CCGCCCGGAG CAGCTGCGCT GTCGGGGCCA 4830
GGCCGGGCTC CCAGTGGATT CGCGGGCACA GACGCCCAGG ACCGCGCTCC CCACGTGGCG GAGGGACTGG 4900
GGACCCGGGC ACCCGTCCTG CCCCTTCACC TTCCAGCTCC GCCTCCTCCG CGCGGACCCC GCCCCGTCCC 4970
GACCCCTCCC GGGTCCCCGG CCCAGCCCCC TCCGGGCCCT CCCAGCCCCT CCCCTTCCTT TCCGCGGCCC 5040
CGCCCTCTCC TCGCGGCGCG AGTTTCAGGC AGCGCTGCGT CCTGCTGCGC ACGTGGGAAG CCCTGGCCCC 5110
GGCCACCCCC GCGATG                                                           5126
```

Fig. 5

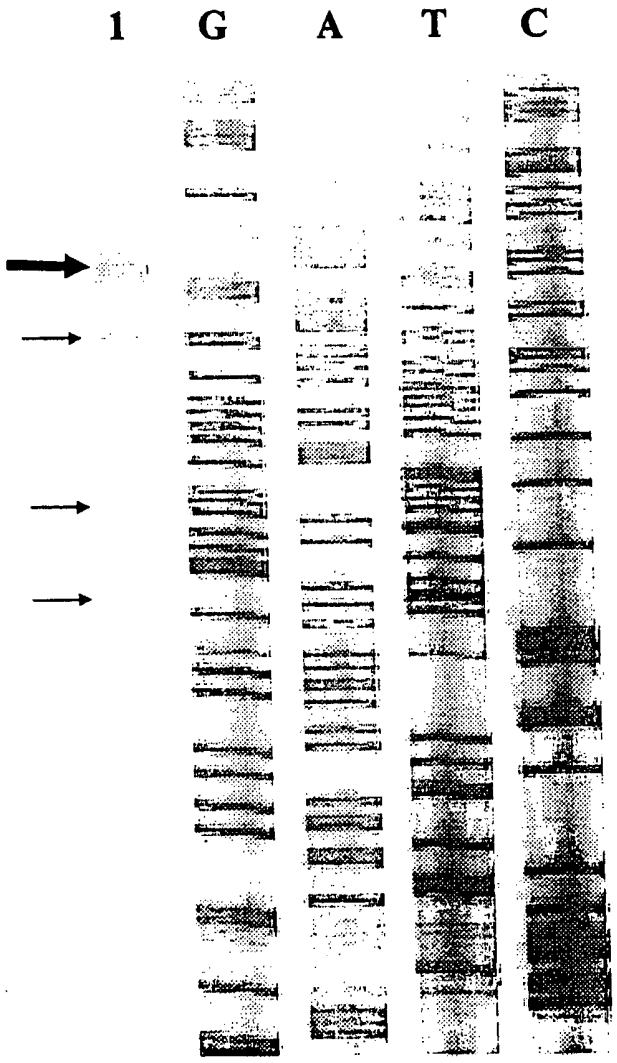


Fig. 6

```
GTTTCAGGCA GCGCTGCGTC CTGCTGCGCA CGTGGGAAGC CCTGGCCCCG GCCACCCCCG CGATGCCGCG   70
CGCTCCCCGC TGCCGAGCCG TGCGCTCCCT GCTGCGCAGC CACTACCGCG AGGTGCTGCC GCTGGCCACG  140
TTCGTGCGGC GCCTGGGGCC CCAGGGCTGG CGGCTGGTGC AGCGCGGGGA CCCGGCGGCT TTCCGCGCGC  210
TGGTGGCCCA GTGCCTGGTG TGCGTGCCCT GGGACGCACG GCCGCCCCCC GCCGCCCCCT CCTTCCGCCA  280
GGTGTCCTGC CTGAAGGAGC TGGTGGCCCG AGTGCTGCAG AGGCTGTGCG AGCGCGGCGC GAAGAACGTG  350
CTGGCCTTCG GCTTCGCGCT GCTGGACGGG GCCCGCGGGG GCCCCCCCGA GGCCTTCACC ACCAGCGTGC  420
GCAGCTACCT GCCCAACACG GTGACCGACG CACTGCGGGG GAGCGGGGCG TGGGGGCTGC TGCTGCGCCG  490
CGTGGGCGAC GACGTGCTGG TTCACCTGCT GGCACGCTGC GCGCTCTTTG TGCTGGTGGC TCCCAGCTGC  560
GCCTACCAGG TGTGCGGGCC GCCGCTGTAC CAGCTCGGCG CTGCCACTCA GGCCCGGCCC CCGCCACACG  630
CTAGTGGACC CCGAAGGCGT CTGGGATGCG AACGGGCCTG GAACCATAGC GTCAGGGAGG CCGGGGTCCC  700
CCTGGGCCTG CCAGCCCCGG GTGCGAGGAG GCGCGGGGGC AGTGCCAGCC GAAGTCTGCC GTTGCCCAAG  770
AGGCCCAGGC GTGGCGCTGC CCCTGAGCCG GAGCGGACGC CCGTTGGGCA GGGGTCCTGG GCCCACCCGG  840
GCAGGACGCG TGGACCGAGT GACCGTGGTT TCTGTGTGGT GTCACCTGCC AGACCCGCCG AAGAAGCCAC  910
CTCTTTGGAG GGTGCGCTCT CTGGCACGCG CCACTCCCAC CCATCCGTGG GCCGCCAGCA CCACGCGGGC  980
CCCCCATCCA CATCGCGGCC ACCACGTCCC TGGGACACGC CTTGTCCCCC GGTGTACGCC GAGACCAAGC 1050
ACTTCCTCTA CTCCTCAGGC GACAAGGAGC AGCTGCGGCC CTCCTTCCTA CTCAGCTCTC TGAGGCCCAG 1120
CCTGACTGGC GCTCGGAGGC TCGTGGAGAC CATCTTTCTG GGTTCCAGGC CCTGGATGCC AGGGACTCCC 1190
CGCAGGTTGC CCCGCCTGCC CCAGCGCTAC TGGCAAATGC GGCCCCTGTT TCTGGAGCTG CTTGGGAACC 1260
ACGCGCAGTG CCCCTACGGG GTGCTCCTCA AGACGCACTG CCCGCTGCGA GCTGCGGTCA CCCCAGCAGC 1330
CGGTGTCTGT GCCCGGGAGA AGCCCCAGGG CTCTGTGGCG GCCCCCGAGG AGGAGGACAC AGACCCCCGT 1400
CGCCTGGTGC AGCTGCTCCG CCAGCACAGC AGCCCCTGGC AGGTGTACGG CTTCGTGCGG GCCTGCCTGC 1470
GCCGGCTGGT GCCCCCAGGC CTCTGGGGCT CCAGGCACAA CGAACGCCGC TTCCTCAGGA ACACCAAGAA 1540
GTTCATCTCC CTGGGGAAGC ATGCCAAGCT CTCGCTGCAG GAGCTGACGT GGAAGATGAG CGTGCGGGAC 1610
TGCGCTTGGC TGCGCAGGAG CCCAGGGGTT GGCTGTGTTC CGGCCGCAGA GCACCGTCTG CGTGAGGAGA 1680
TCCTGGCCAA GTTCCTGCAC TGGCTGATGA GTGTGTACGT CGTCGAGCTG CTCAGGTCTT TCTTTTATGT 1750
CACGGAGACC ACGTTTCAAA AGAACAGGCT CTTTTTCTAC CGGAAGAGTG TCTGGAGCAA GTTGCAAAGC 1820
ATTGGAATCA GACAGCACTT GAAGAGGGTG CAGCTGCGGG AGCTGTCGGA AGCAGAGGTC AGGCAGCATC 1890
GGGAAGCCAG GCCCGCCCTG CTGACGTCCA GACTCCGCTT CATCCCCAAG CCTGACGGGC TGCGGCCGAT 1960
TGTGAACATG GACTACGTCG TGGGAGCCAG AACGTTCCGC AGAGAAAAGA GGGCCGAGCG TCTCACCTCG 2030
AGGGTGAAGG CACTGTTCAG CGTGCTCAAC TACGAGCGGG CGCGGCGCCC CGGCCTCCTG GGCGCCTCTG 2100
TGCTGGGCCT GGACGATATC CACAGGGCCT GGCGCACCTT CGTGCTGCGT GTGCGGGCCC AGGACCCGCC 2170
GCCTGAGCTG TACTTTGTCA AGGTGGATGT GACGGGCGCG TACGACACCA TCCCCCAGGA CAGGCTCACG 2240
GAGGTCATCG CCAGCATCAT CAAACCCCAG AACACGTACT GCGTGCGTCG GTATGCCGTG GTCCAGAAGG 2310
CCGCCCATGG GCACGTCCGC AAGGCCTTCA AGAGCCACGT CTCTACCTTG ACAGACCTCC AGCCGTACAT 2380
GCGACAGTTC GTGGCTCACC TGCAGGAGAC CAGCCCGCTG AGGGATGCCG TCGTCATCGA GCAGAGCTCC 2450
TCCCTGAATG AGGCCAGCAG TGGCCTCTTC GACGTCTTCC TACGCTTCAT GTGCCACCAC GCCGTGCGCA 2520
TCAGGGGCAA GTCCTACGTC CAGTGCCAGG GGATCCCGCA GGGCTCCATC CTCTCCACGC TGCTCTGCAG 2590
CCTGTGCTAC GGCGACATGG AGAACAAGCT GTTTGCGGGG ATTCGGCGGG ACGGGCTGCT CCTGCGTTTG 2660
GTGGATGATT TCTTGTTGGT GACACCTCAC CTCACCCACG CGAAAACCTT CCTCAGGACC CTGGTCCGAG 2730
GTGTCCCTGA GTATGGCTGC GTGGTGAACT TGCGGAAGAC AGTGGTGAAC TTCCCTGTAG AAGACGAGGC 2800
CCTGGGTGGC ACGGCTTTTG TTCAGATGCC GGCCCACGGC CTATTCCCCT GGTGCGGCCT GCTGCTGGAT 2870
ACCCGGACCC TGGAGGTGCA GAGCGACTAC TCCAGCTATG CCCGGACCTC CATCAGAGCC AGTCTCACCT 2940
TCAACCGCGG CTTCAAGGCT GGGAGGAACA TGCGTCGCAA ACTCTTTGGG GTCTTGCGGC TGAAGTGTCA 3010
CAGCCTGTTT CTGGATTTGC AGGTGAACAG CCTCCAGACG GTGTGCACCA ACATCTACAA GATCCTCCTG 3080
CTGCAGGCGT ACAGGTTTCA CGCATGTGTG CTGCAGCTCC CATTTCATCA GCAAGTTTGG AAGAACCCCA 3150
CATTTTTCCT GCGCGTCATC TCTGACACGG CCTCCCTCTG CTACTCCATC CTGAAAGCCA AGAACGCAGG 3220
GATGTCGCTG GGGGCCAAGG GCGCCGCCGG CCCTCTGCCC TCCGAGGCCG TGCAGTGGCT GTGCCACCAA 3290
GCATTCCTGC TCAAGCTGAC TCGACACCGT GTCACCTACG TGCCACTCCT GGGGTCACTC AGGACAGCCC 3360
AGACGCAGCT GAGTCGGAAG CTCCCGGGGA CGACGCTGAC TGCCCTGGAG GCCGCAGCCA ACCCGGCACT 3430
GCCCTCAGAC TTCAAGACCA TCCTGGACTG ATGGCCACCC GCCCACAGCC AGGCCGAGAG CAGACACCAG 3500
CAGCCCTGTC ACGCCGGGCT CTACGTCCCA GGGAGGGAGG GGCGGCCCAC ACCCAGGCCC GCACCGCTGG 3570
GAGTCTGAGG CCTGAGTGAG TGTTTGGCCG AGGCCTGCAT GTCCGGCTGA AGGCTGAGTG TCCGGCTGAG 3640
GCCTGAGCGA GTGTCCAGCC AAGGGCTGAG TGTCCAGCAC ACCTGCCGTC TTCACTTCCC CACAGGCTGG 3710
CGCTCGGCTC CACCCCAGGG CCAGCTTTTC CTCACCAGGA GCCCGGCTTC CACTCCCCAC ATAGGAATAG 3780
TCCATCCCCA GATTCGCCAT TGTTCACCCC TCGCCCTGCC CTCCTTTGCC TTCCACCCCC ACCATCCAGG 3850
TGGAGACCCT GAGAAGGACC CTGGGAGCTC TGGGAATTTG GAGTGACCAA AGGTGTGCCC TGTACACAGG 3920
CGAGGACCCT GCACCTGGAT GGGGGTCCCT GTGGGTCAAA TTGGGGGGAG GTGCTGTGGG AGTAAAATAC 3990
TGAATATATG AGTTTTTCAG TTTTGAAAAA AAAAAAAAAA AAAAAAAAAA AA                     4042
```

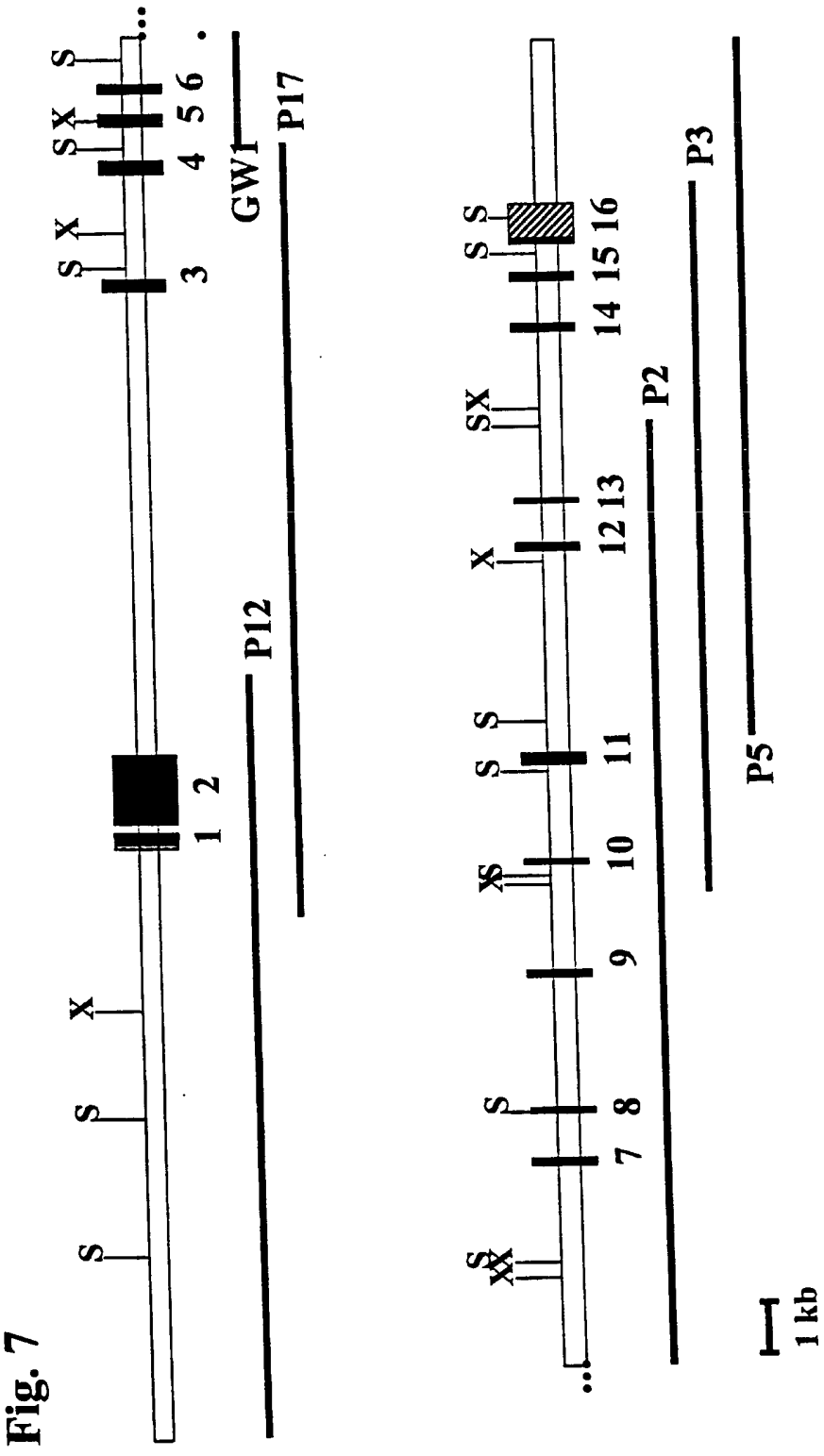

S
S X
S X
S
X
S
S
X
S
S
X
S
S
X
S
S
S
X
S
S
1 2
3
4 5 6
GW1
P17
P12
12 13
14 15 16
P2
P3
P5
1 kb

Fig. 7

Fig. 8A

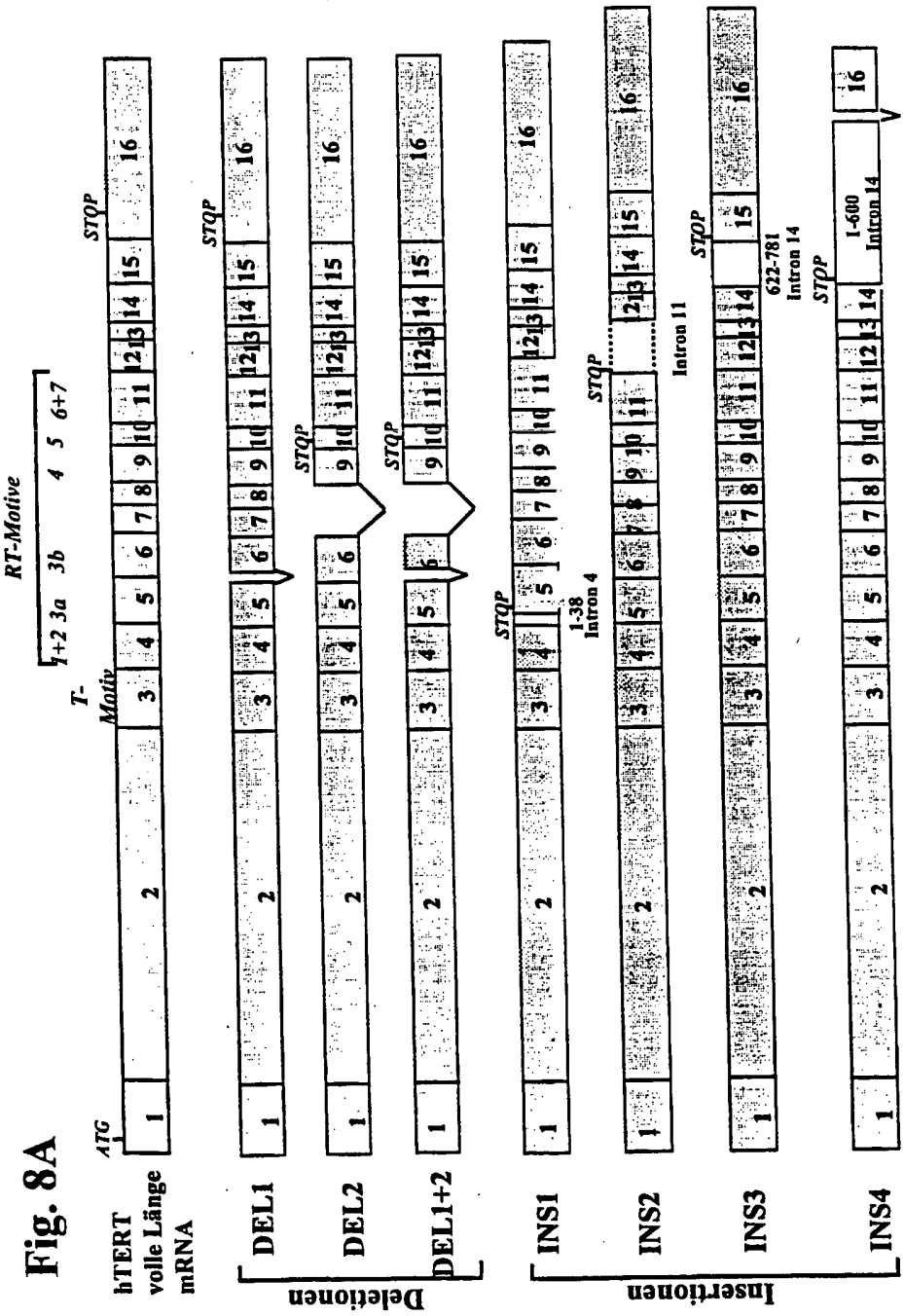
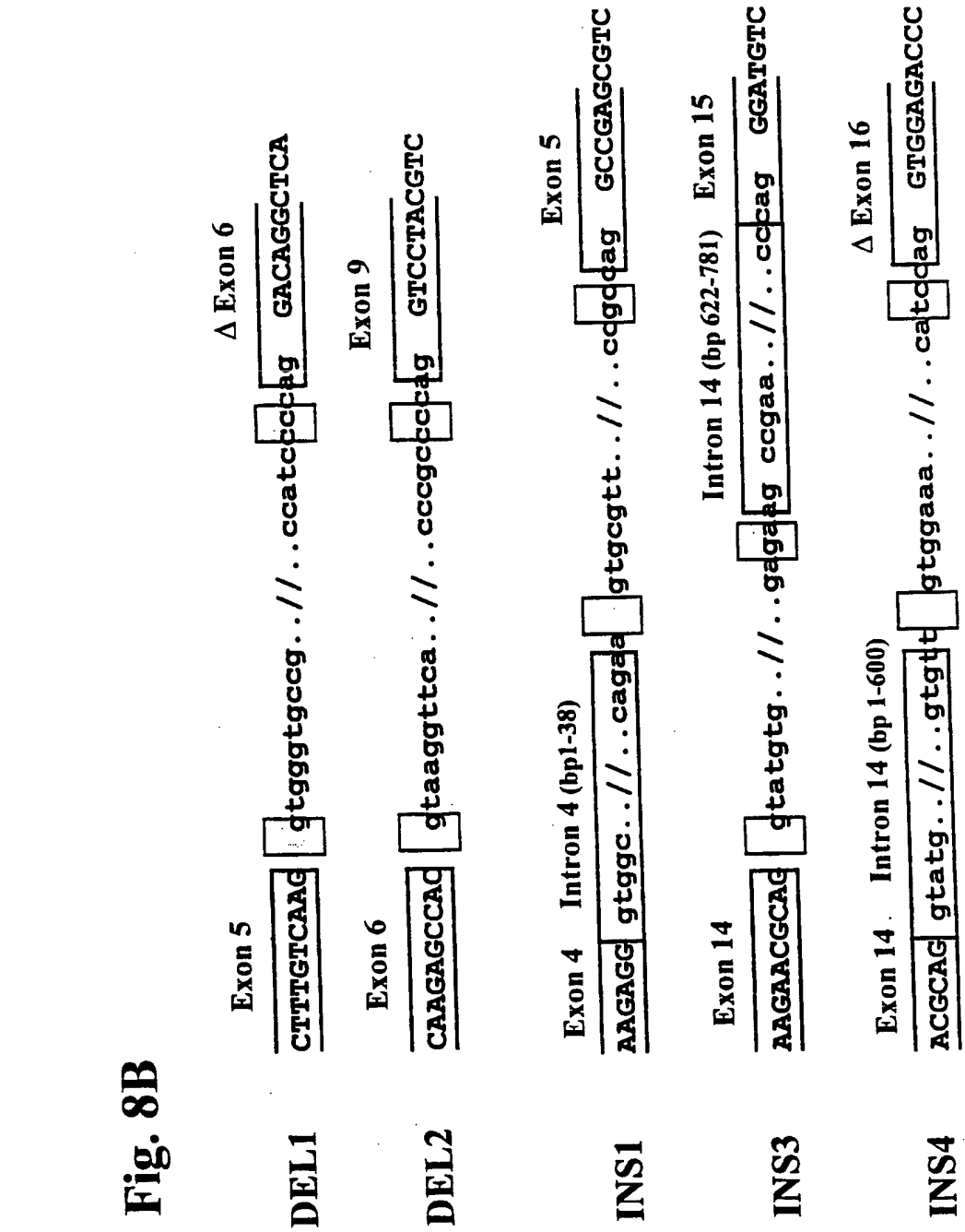

RT-Motive
T-Motiv
1+2 3a 3b 4 5 6+7
ATG
STOP
hTERT volle Länge mRNA
DEL1
DEL2
DEL1+2
Deletionen
INS1
INS2
INS3
INS4
Insertionen
1-38 Intron 4
Intron 11
622-781 Intron 14
1-600 Intron 14

Fig. 8B


Δ Exon 6
Exon 5
DEL1
CTTTGTCAAG gtgggtgccg..//..ccatcccag GACAGGCTCA
Exon 9
Exon 6
DEL2
CAAGAGCCAG gtaaggttca..//..cccgcccag GTCCTACGTC
Exon 5
Exon 4 Intron 4 (bp1-38)
INS1
AAGAGG gtggc..//..caga gtgcgtt..//..cgg ag GCCGAGCGTC
Intron 14 (bp 622-781) Exon 15
Exon 14
INS3
AAGAACGCAG gtatgtg..//..gaga ag ccgaa..//..ccc ag GGATGTC
Δ Exon 16
Exon 14 Intron 14 (bp 1-600)
INS4
ACGCAG gtatg..//..gtgtt gtggaaa..//..catc ag GTGGAGACCC

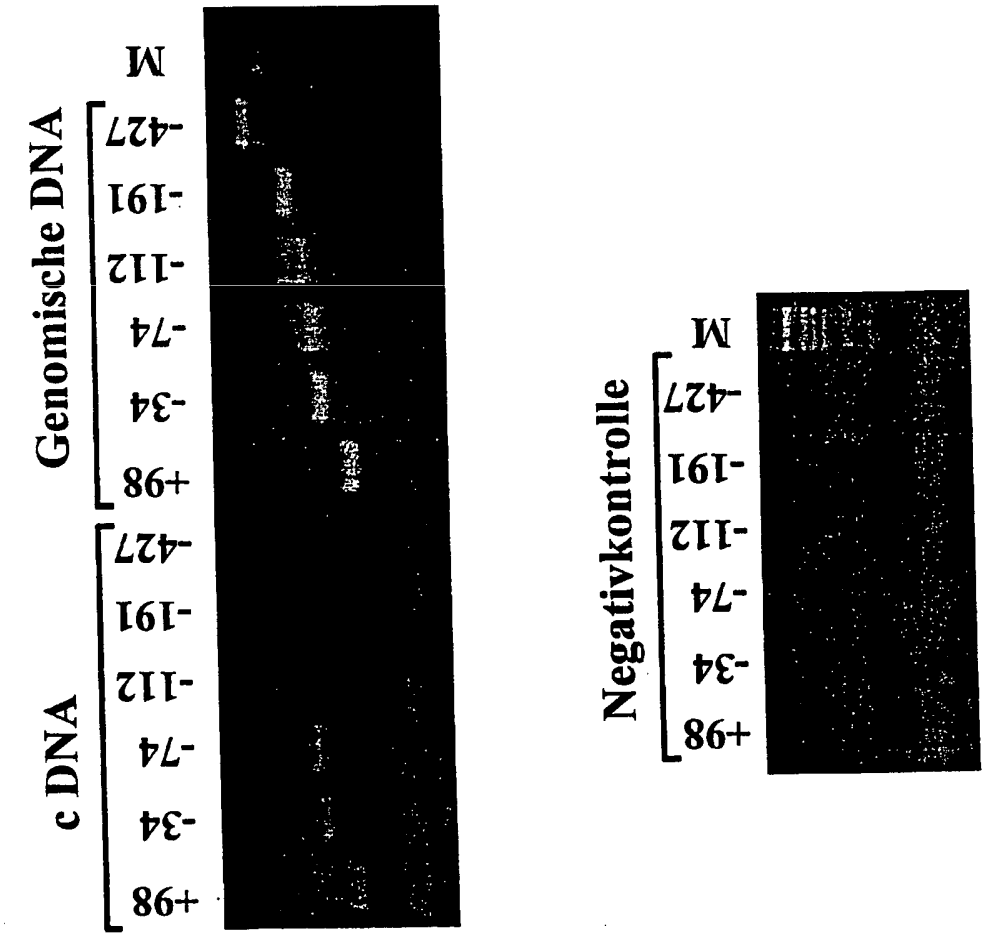

Genomische DNA
M
-427
-191
-112
-74
-34
+98
c DNA
-427
-191
-112
-74
-34
+98
Negativkontrolle
M
-427
-191
-112
-74
-34
+98

Fig. 9

Fig. 10

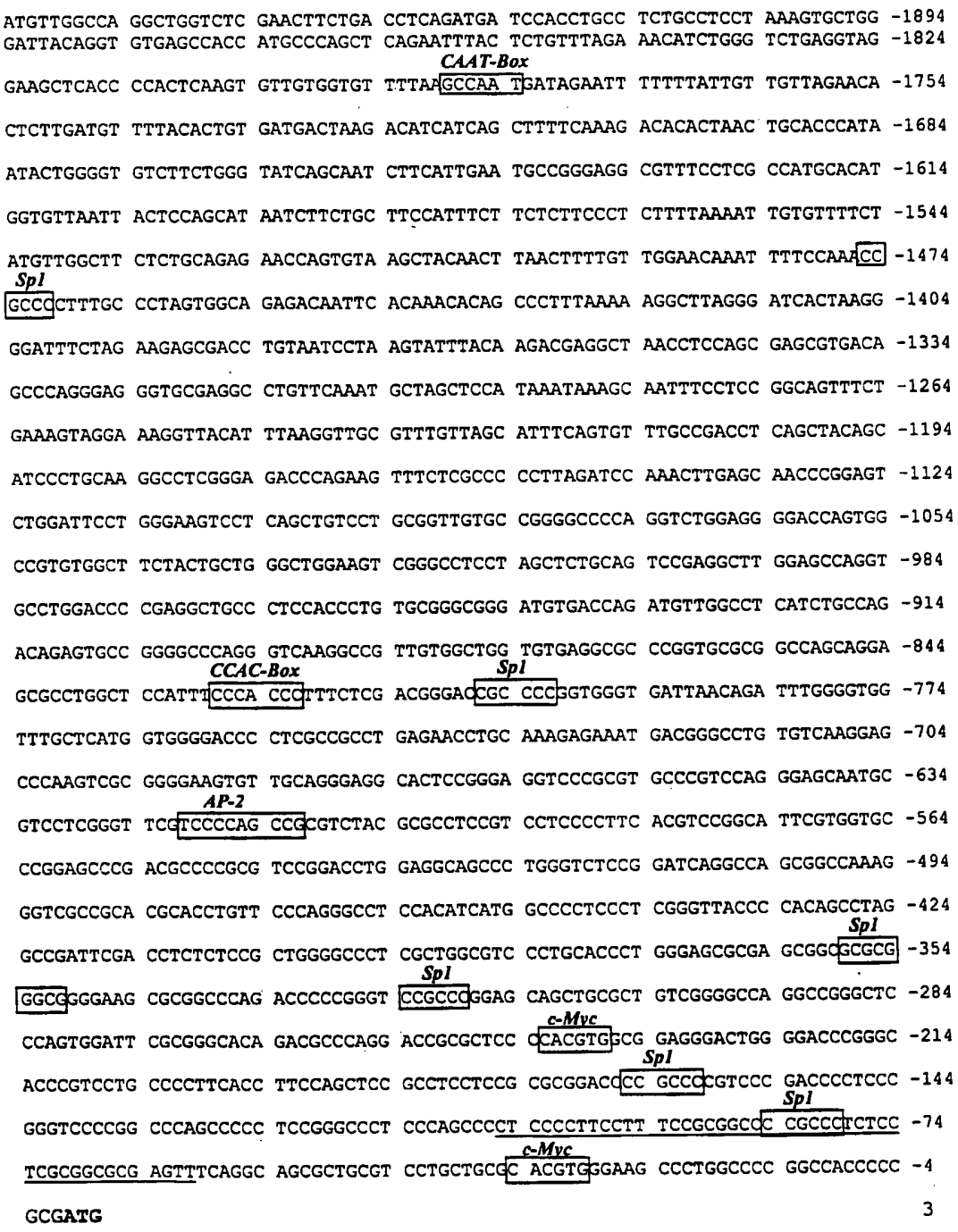
```
ACTTGAGCCC AAGAGTTCAA GGCTACGGTG AGCCATGATT GCAACACCAC ACGCCAGCCT TGGTGACAGA -11204
ATGAGACCCT GTCTCAAAAA AAAAAAAAAA AATTGAAATA ATATAAAGCA TCTTCTCTGG CCACAGTGGA -11134
ACAAAACCAG AAATCAACAA CAAGAGGAAT TTTGAAAACT ATACAAACAC ATGAAAATTA AACAATATAC -11064
TTCTGAATGA CCAGTGAGTC AATGAAGAAA TTAAAAAGGA AATTGAAAAA TTTATTTAAG CAAATGATAA -10994
CGGAAACATA ACCTCTCAAA ACCCACGGTA TACAGCAAAA GCAGTGCTAA GAAGGAAGTT TATAGCTATA -10924
AGCAGCTACA TCAAAAAAGT AGAAAAGCCA GGCGCAGTGG CTCATGCCTG TAATCCCAGC ACTTTGGGAG -10854
GCCAAGGCGG GCAGATCGCC TGAGGTCAGG AGTTCGAGAC CAGCCTGACC AACACAGAGA AACCTTGTCG -10784
CTACTAAAAA TACAAAATTA GCTGGGCATG GTGGCACATG CCTGTAATCC CAGCTACTCG GGAGGCTGAG -10714
GCAGGATAAC CGCTTGAACC CAGGAGGTGG AGGTTGCGGT GAGCCGGGAT TGCGCCATTG GACTCCAGCC -10644
TGGGTAACAA GAGTGAAACC CTGTCTCAAG AAAAAAAAAA AAGTAGAAAA ACTTAAAAAT ACAACCTAAT -10574
GATGCACCTT AAAGAACTAG AAAAGCAAGA GCAAACTAAA CCTAAAATTG GTAAAAGAAA AGAAATAATA -10504
AAGATCAGAG CAGAAATAAA TGAAACTGAA AGATAACAAT ACAAAAGATC AACAAAATTA AAAGTTGGTT -10434
TTTTGAAAAG ATAAACAAAA TTGACAAACC TTTGCCCAGA CTAAGAAAAA AGGAAAGAAG ACCTAAATAA -10364
ATAAAGTCAG AGATGAAAAA AGAGACATTA CAACTGATAC CACAGAAATT CAAAGGATCA CTAGAGGCTA -10294
CTATGAGCAA CTGTACACTA ATAAATTGAA AAACCTAGAA AAAATAGATA AATTCCTAGA TGCATACAAC -10224
CTACCAAGAT TGAACCATGA AGAAATCCAA AGCCCAAACA GACCAATAAC AATAATGGGA TTAAAGCCAT -10154
AATAAAAAGT CTCCTAGCAA AGAGAAGCCC AGGACCCAAT GGCTTCCCTG CTGGATTTTA CCAATCATTT -10084
AAAGAAGAAT GAATTCCAAT CCTACTCAAA CTATTCTGAA AAATAGAGGA AAGAATACTT CCAAACTCAT -10014
TCTACATGGC CAGTATTACC CTGATTCCAA AACCAGACAA AAACACATCA AAAACAAACA AACAAAAAAA -9944
CAGAAAGAAA GAAAACTACA GGCCAATATC CCTGATGAAT ACTGATACAA AAATCCTCAA CAAAACACTA -9874
GCAAACCAAA TTAAACAACA CCTTCGAAAG ATCATTCATT GTGATCAAGT GGGATTTATT CCAGGGATGG -9804
AAGGATGGTT CAACATATGC AAATCAATCA ATGTGATACA TCATCCCAAC AAAATGAAGT ACAAAAACTA -9734
TATGATTATT TCACTTTATG CAGAAAAAGC ATTTGATAAA ATTCTGCACC CTTCATGATA AAAACCCTCA -9664
AAAAACCAGG TATACAAGAA ACATACAGGC CAGGCACAGT GGCTCACACC TGCGATCCCA GCACTCTGGG -9594
AGGCCAAGGT GGGATGATTG CTTGGGCCCA GGAGTTTGAG ACTAGCCTGG GCAACAAAAT GAGACCTGGT -9524
CTACAAAAAA CTTTTTTAAA AAATTAGCCA GGCATGATGG CATATGCCTG TAGTCCCAGC TAGTCTGGAG -9454
GCTGAGGTGG GAGAATCACT TAAGCCTAGG AGGTCGAGGC TGCAGTGAGC CATGAACATG TCACTGTACT -9384
CCAGCCTAGA CAACAGAACA AGACCCCACT GAATAAGAAG AAGGAGAAGG AGAAGGGAGA AGGGAGGGAG -9314
AAGGGAGGAG GAGGAGAAGG AGGAGGTGGA GGAGAAGTGG AAGGGGAAGG GGAAGGGAAA GAGGAAGAAG -9244
AAGAAACATA TTTCAACATA ATAAAAGCCC TATATGACAG ACCGAGGTAG TATTATGAGG AAAAACTGAA -9174
AGCCTTTCCT CTAAGATCTG GAAAATGACA AGGGCCCACT TTCACCACTG TGATTCAACA TAGTACTAGA -9104
AGTCCTAGCT AGAGCAATCA GATAAGAGAA AGAAATAAAA GGCATCCAAA CTGGAAAGGA AGAAGTCAAA -9034
TTATCCTGTT TGCAGATGAT ATGATCTTAT ATCTGGAAAA GACTTAAGAC ACCACTAAAA AACTATTAGA -8964
GCTGAAATTT GGTACAGCAG GATACAAAAT CAATGTACAA AAATCAGTAG TATTTCTATA TTCCAACAGC -8894
AAACAATCTG AAAAAGAAAC CAAAAAAGCA GCTACAAATA AAATTAAACA GCTAGGAATT AACCAAAGAA -8824
GTGAAAGATC TCTACAATGA AAACTATAAA ATGTTGATAA AAGAAATTGA AGAGGGCACA AAAAAAGAAA -8754
AGATATTCCA TGTTCATAGA TTGGAAGAAT AAATACTGTT AAAATGTCCA TACTACCCAA AGCAATTTAC -8684
AAATTCAATG CAATCCCTAT TAAAATACTA ATGACGTTCT TCACAGAAAT AGAAGAAACA ATTCTAAGAT -8614
TTGTACAGAA CCACAAAAGA CCCAGAATAG CCAAAGCTAT CCTGACCAAA AAGAACAAAA CTGGAAGCAT -8544
CACATTACCT GACTTCAAAT TATACTACAA AGCTATAGTA ACCCAAACTA CATGGTACTG GCATAAAAAC -8474
AGATGAGACA TGGACCAGAG GAACAGAATA GAGAATCCAG AAACAAATCC ATGCATCTAC AGTGAACTCA -8404
TTTTTGACAA AGGTGCCAAG AACATACTTT GGGGAAAAGA TAATCTCTTC AATAAATGGT GCTGGAGGAA -8334
CTGGATATCC ATATGCAAAA TAACAATACT AGAACTCTGT CTCTCACCAT ATACAAAAGC AAATCAAAAT -8264
GGATGAAAGG CTTAAATCTA AAACCTCAAA CTTTGCAACT ACTAAAAGAA AACACCGGAG AAACTCTCCA -8194
GGACATTGGA GTGGGCAAAG ACTTCTTGAG TAATTCCCTG CAGGCACAGG CAACCAAAGC AAAAACAGAC -8124
AAATGGGATC ATATCAAGTT AAAAAGCTTC TGCCCAGCAA AGGAAACAAT CAACAAAGAG AAGAGACAAC -8054
CCACAGAATG GGAGAATATA TTTGCAAACT ATTCATCTAA CAAGGAATTA ATAACCAGTA TATATAAGGA -7984
GCTCAAACTA CTCTATAAGA AAAACACCTA ATAAGCTGAT TTTCAAAAAT AAGCAAAAGA TCTGGGTAGA -7914
CATTTCTCAA AATAAGTCAT ACAAATGGCA AACAGGCATC TGAAAATGTG CTCAACACCA CTGATCATCA -7844
GAGAAATGCA AATCAAAACT ACTATGAGAG ATCATCTCAT CCCAGTTAAA ATGGCTTTTA TTCAAAAGAC -7774
AGGCAATAAC AAATGCCAGT GAGGATGTGG ATAAAAGGAA ACCCTTGGAC ACTGTTGGTG GGAATGGAAA -7704
TTGCTACCAC TATGGAGAAC AGTTTGAAAG TTCCTCAAAA AACTAAAAAT AAAGCTACCA TACAGCAATC -7634
CCATTGCTAG GTATATACTC CAAAAAAGGG AATCAGTGTA TCAACAAGCT ATCTCCACTC CCACATTTAC -7564
TGCAGCACTG TTCATAGCAG CCAAGGTTTG GAAGCAACCT CAGTGTCCAT CAACAGACGA ATGGAAAAAG -7494
AAAATGTGGT GCACATACAC AATGGAGTAC TACGCAGCCA TAAAAAAGAA TGAGATCCTG TCAGTTGCAA -7424
CAGCATGGGG GGCACTGGTC AGTATGTTAA GTGAAATAAG CCAGGCACAG AAAGACAAAC TTTTCATGTT -7354
CTCCCTTACT TGTGGGAGCA AAAATTAAAA CAATTGACAT AGAAATAGAG GAGAATGGTG GTTCTAGAGG -7284
GGTGGGGGAC AGGGTGACTA GAGTCAACAA TAATTTATTG TATGTTTTAA AATAACTAAA AGAGTATAAT -7214
TGGGTTGTTT GTAACACAAA GAAAGGATAA ATGCTTGAAG GTGACAGATA CCCCATTTAC CCTGATGTGA -7144
TTATTACACA TTGTATGCCT GTATCAAAAT ATCTCATGTA TGCTATAGAT ATAAACCCTA CTATATTAAA -7074
AATTAAAATT TTAATGGCCA GGCACGGTGG CTCATGTCCG TAATCCCAGC ACTTTGGGAG GCCGAGGCGG -7004
GTGGATCACC TGAGGTCAGG AGTTTGAAAC CAGTCTGGCC ACCATGATGA AACCCTGTCT CTACTAAAGA -6934
TACAAAAATT AGCCAGGCGT GGTGGCACAT ACCTGTAGTC CCAACTACTC AGGAGGCTGA GACAGGAGAA -6864
TTGCTTGAAC CTGGGAGGCG GAGGTTGCAG TGAGCCGAGA TCATGCCACT GCACTGCAGC CTGGGTGACA -6794
GAGCAAGACT CCATCTCAAA ACAAAAACAA AAAAAGAAG ATTAAAATTG TAATTTTTAT GTACCGTATA -6724
AATATATACT CTACTATATT AGAAGTTAAA AATTAAAACA ATTATAAAAG GTAATTAACC ACTTAATCTA -6654
AAATAAGAAC AATGTATGTG GGGTTTCTAG CTTCTGAAGA AGTAAAAGTT ATGGCCACGA TGGCAGAAAT -6584
```

Fig. 10

GTGAGGAGGG AACAGTGGAA GTTACTGTTG TTAGACGCTC ATACTCTCTG TAAGTGACTT AATTTTAACC -6514
AAAGACAGGC TGGGAGAAGT TAAAGAGGCA TTCTATAAGC CCTAAAACAA CTGCTAATAA TGGTGAAAGG -6444
TAATCTCTAT TAATTACCAA TAATTACAGA TATCTCTAAA ATCGAGCTGC AGAATTGGCA CGTCTGATCA -6374
CACCGTCCTC TCATTCACGG TGCTTTTTTT CTTGTGTGCT TGGAGATTTT CGATTGTGTG TTCGTGTTTG -6304
GTTAAACTTA ATCTGTATGA ATCCTGAAAC GAAAAATGGT GGTGATTTCC TCCAGAAGAA TTAGAGTACC -6234
TGGCAGGAAG CAGGTGGCTC TGTGGACCTG AGCCACTTCA ATCTTCAAGG GTCTCTGGCC AAGACCCAGG -6164
TGCAAGGCAG AGGCCTGATG ACCCGAGGAC AGGAAAGCTC GGATGGGAAG GGGCGATGAG AAGCCTGCCT -6094
CGTTGGTGAG CAGCGCATGA AGTGCCCTTA TTTACGCTTT GCAAAGATTG CTCTGGATAC CATCTGGAAA -6024
AGGCGGCCAG CGGGAATGCA AGGAGTCAGA AGCCTCCTGC TCAAACCCAG GCCAGCAGCT ATGGCGCCCA -5954
CCCGGGCGTG TGCCAGAGGG AGAGGAGTCA AGGCACCTCG AAGTATGGCT TAAATCTTTT TTTCACCTGA -5884
AGCAGTGACC AAGGTGTATT CTGAGGGAAG CTTGAGTTAG GTGCCTTCTT TAAAACAGAA AGTCATGGAA -5814
GCACCCTTCT CAAGGGAAAA CCAGACGCCC GCTCTGCGGT CATTTACCTC TTTCCTCTCT CCCTCTCTTG -5744
CCCTCGCGGT TTCTGATCGG GACAGAGTGA CCCCCGTGGA GCTTCTCCGA GCCCGTGCTG AGGACCCTCT -5674
TGCAAAGGGC TCCACAGACC CCCGCCCTGG AGAGAGGAGT CTGAGCCTGG CTTAATAACA AACTGGGATG -5604
TGGCTGGGGG CGGACAGCGA CGGCGGGATT CAAAGACTTA ATTCCATGAG TAAATTCAAC CTTTCCACAT -5534
CCGAATGGAT TTGGATTTTA TCTTAATATT TTCTTAAATT TCATCAAATA ACATTCAGGA CTGCAGAAAT -5464
CCAAAGGCGT AAAACAGGAA CTGAGCTATG TTTGCCAAGG TCCAAGGACT TAATAACCAT GTTCAGAGGG -5394
ATTTTTCGCC CTAAGTACTT TTTATTGGTT TTCATAAGGT GGCTTAGGGT GCAAGGGAAA GTACACGAGG -5324
AGAGGCCTGG GCGGCAGGGC TATGAGCACG GCAGGGCCAC CGGGGAGAGA GTCCCCGGCC TGGGAGGCTG -5254
ACAGCAGGAC CACTGACCGT CCTCCCTGGG AGCTGCCACA TTGGGCAACG CGAAGGCGGC CACGCTGCGT -5184
GTGACTCAGG ACCCCATACC GGCTTCCTGG GCCCACCCAC ACTAACCCAG GAAGTCACGG AGCTCTGAAC -5114
CCGTGGAAAC GAACATGACC CTTGCCTGCC TGCTTCCCTG GGTGGGTCAA GGGTAATGAA GTGGTGTGCA -5044
GGAAATGGCC ATGTAAATTA CACGACTCTG CTGATGGGGA CCGTTCCTTC CATCATTATT CATCTTCACC -4974
CCCAAGGACT GAATGATTCC AGCAACTTCT TCGGGTGTGA CAAGCCATGA CAAAACTCAG TACAAACACC -4904
ACTCTTTTAC TAGGCCCACA GAGCACGGSC CACACCCCTG ATATATTAAG AGTCCAGGAG AGATGAGGCT -4834
GCTTTCAGCC ACCAGGCTGG GGTGACAACA GCGGCTGAAC AGTCTGTTCC TCTAGACTAG TAGACCCTGG -4764
CAGGCACTCC CCCAGATTCT AGGGCCTGGT TGCTGCTTCC CGAGGGCGCC ATCTGCCCTG GAGACTCAGC -4694
CTGGGGTGCC ACACTGAGGC CAGCCCTGTC TCCACACCCT CCGCCTCCAG GCCTCAGCTT CTCCAGCAGC -4624
TTCCTAAACC CTGGGTGGGC CGTGTTCCAG CGCTACTGTC TCACCTGTCC CACTGTGTCT TGTCTCAGCG -4554
ACGTAGCTCG CACGGTTCCT CCTCACATGG GGTGTCTGTC TCCTTCCCCA ACACTCACAT GCGTTGAAGG -4484
GAGGAGATTC TGCGCCTCCC AGACTGGCTC CTCTGAGCCT GAACCTGGCT CGTGGCCCCC GATGCAGGTT -4414
CCTGGCGTCC GGCTGCACGC TGACCTCCAT TTCCAGGCGC TCCCCGTCTC CTGTCATCTG CCGGGGCCTG -4344
CCGGTGTGTT CTTCTGTTTC TGTGCTCCTT TCCACGTCCA GCTGCGTGTG TCTCTGCCCG CTAGGGTCTC -4274
GGGGTTTTTA TAGGCATAGG ACGGGGGCGT GGTGGGCCAG GGCGCTCTTG GGAAATGCAA CATTTGGGTG -4204
TGAAAGTAGG AGTGCCTGTC CTCACCTAGG TCCACGGGCA CAGGCCTGGG GATGGAGCCC CCGCCAGGGA -4134
CCCGCCCTTC TCTGCCCAGC ACTTTCCTGC CCCCCTCCCT CTGGAACACA GAGTGGCAGT TTCCACAAGC -4064
ACTAAGCATC CTCTTCCCAA AAGACCCAGC ATTGGCACCC CTGGACATTT GCCCCACAGC CCTGGGAATT -3994

c-Myc
CACGTGACTA CGCACATCAT GTACACACTC CCGTCCACGA CCGACCCCCG CTGTTTTATT TTAATAGCTA -3924
CAAAGCAGGG AAATCCCTGC TAAAATGTCC TTTAACAAAC TGGTTAAACA AACGGGTCCA TCCGCACGGT -3854
GGACAGTTCC TCACAGTGAA GAGGAACATG CCGTTTATAA AGCCTGCAGG CATCTCAAGG GAATTACGCT -3784
GAGTCAAAAC TGCCACCTCC ATGGGATACG TACGCAACAT GCTCAAAAAG AAAGAATTTC ACCCCATGGC -3714
AGGGGAGTGG TTAGGGGGGT TAAGGACGGT GGGGGCGGCA GCTGGGGGCT ACTGCACGCA CCTTTTACTA -3644
AAGCCAGTTT CCTGGTTCTG ATGGTATTGG CTCAGTTATG GGAGACTAAC CATAGGGGAG TGGGGATGGG -3574
GGAACCCGGA GGCTGTGCCA TCTTTGCCAT GCCCGAGTGT CCTGGGCAGG ATAATGCTCT AGAGATGCCC -3504
ACGTCCTGAT TCCCCCAAAC CTGTGGACAG AACCCGCCCG GCCCCAGGGC CTTTGCAGGT GTGATCTCCG -3434
TGAGGACCCT GAGGTCTGGG ATCCTTCGGG ACTACCTGCA GGCCCGAAAA GTAATCCAGG GGTTCTGGGA -3364
AGAGGCGGGC AGGAGGGTCA GAGGGGGGCA GCCTCAGGAC GATGGAGGCA GTCAGTCTGA GGCTGAAAAG -3294
GGAGGGAGGG CCTCGAGCCC AGGCCTGCAA GCGCCTCCAG AAGCTGGAAA AAGCGGGGAA GGGACCCTCC -3224
ACGGAGCCTG CAGCAGGAAG GCACGGCTGG CCCTTAGCCC ACCAGGGCCC ATCGTGGACC TCCGGCCTCC -3154
GTGCCATAGG AGGGCACTCG CGCTGCCCTT CTAGCATGAA GTGTGTGGGG ATTTGCAGAA GCAACAGGAA -3084
ACCCATGCAC TGTGAATCTA GGATTATTTC AAAACAAAGG TTTACAGAAA CATCCAAGGA CAGGGCTGAA -3014
GTGCCTCCGG GCAAGGGCAG GGCAGGCACG AGTGATTTTA TTTAGCTATT TTATTTTATT TACTTACTTT -2944
CTGAGACAGA GTTATGCTCT TGTTGCCCAG GCTGGAGTGC AGCGGCATGA TCTTGGCTCA CTGCAACCTC -2874
CGTCTCCTGG GTTCAAGCAA TTCTCGTGCC TCAGCCTCCC AAGTAGCTGG GATTTCAGGC GTGCACCACC -2804
ACACCCGGCT AATTTTGTAT TTTTAGTAGA GATGGGCTTT CACCATGTTG GTCAAGCTGA TCTCAAAATC -2734
CTGACCTCAG GTGATCCGCC CACCTCAGCC TCCCAAAGTG CTGGGATTAC AGGCATGAGC CACTGCACCT -2664
GGCCTATTTA ACCATTTTAA AACTTCCCTG GGCTCAAGTC ACACCCACTG GTAAGGAGTT CATGGAGTTC -2594
AATTTCCCCT TTACTCAGGA GTTACCCTCC TTTGATATTT TCTGTAATTC TTCGTAGACT GGGGATACAC -2524
CGTCTCTTGA CATATTCACA GTTTCTGTGA CCACCTGTTA TCCCATGGGA CCCACTGCAG GGGCAGCTGG -2454
GAGGCTGCAG GCTTCAGGTC CCAGTGGGGT TGCCATCTGC CAGTAGAAAC CTGATGTAGA ATCAGGGCGC -2384
AAGTGTGGAC ACTGTCCTGA ATCTCAATGT CTCAGTGTGT GCTGAAACAT GTAGAAATTA AAGTCCATCC -2314
CTCCTACTCT ACTGGGATTG AGCCCCTTCC CTATCCCCCC CCAGGGGCAG AGGAGTTCCT CTCACTCCTG -2244
TGGAGGAAGG AATGATACTT TGTTATTTTT CACTGCTGGT ACTGAATCCA CTGTTTCATT TGTTGGTTTG -2174
TTTGTTTTGT TTTGAGAGGC GGTTTCACTC TTGTTGCTCA GGCTGGAGGG AGTGCAATGG CGCGATCTTG -2104
GCTTACTGCA GCCTCTGCCT CCCAGGTTCA AGTGATTCTC CTGCTTCCGC CTCCCATTTG GCTGGGATTA -2034
CAGGCACCCG CCACCATGCC CAGCTAATTT TTTGTATTTT TAGTAGAGAC GGGGGTGGGT GGGGTTCACC -1964

Fig. 10


ATGTTGGCCA GGCTGGTCTC GAACTTCTGA CCTCAGATGA TCCACCTGCC TCTGCCTCCT AAAGTGCTGG -1894
GATTACAGGT GTGAGCCACC ATGCCCAGCT CAGAATTTAC TCTGTTTAGA AACATCTGGG TCTGAGGTAG -1824
CAAT-Box
GAAGCTCACC CCACTCAAGT GTTGTGGTGT TTTAAGCCAA TGATAGAATT TTTTTATTGT TGTTAGAACA -1754
CTCTTGATGT TTTACACTGT GATGACTAAG ACATCATCAG CTTTTCAAAG ACACACTAAC TGCACCCATA -1684
ATACTGGGGT GTCTTCTGGG TATCAGCAAT CTTCATTGAA TGCCGGGAGG CGTTTCCTCG CCATGCACAT -1614
GGTGTTAATT ACTCCAGCAT AATCTTCTGC TTCCATTTCT TCTCTTCCCT CTTTTAAAAT TGTGTTTTCT -1544
ATGTTGGCTT CTCTGCAGAG AACCAGTGTA AGCTACAACT TAACTTTTGT TGGAACAAAT TTTCCAAACC -1474
Sp1
GCCCCTTTGC CCTAGTGGCA GAGACAATTC ACAAACACAG CCCTTTAAAA AGGCTTAGGG ATCACTAAGG -1404
GGATTTCTAG AAGAGCGACC TGTAATCCTA AGTATTTACA AGACGAGGCT AACCTCCAGC GAGCGTGACA -1334
GCCCAGGGAG GGTGCGAGGC CTGTTCAAAT GCTAGCTCCA TAAATAAAGC AATTTCCTCC GGCAGTTTCT -1264
GAAAGTAGGA AAGGTTACAT TTAAGGTTGC GTTTGTTAGC ATTTCAGTGT TTGCCGACCT CAGCTACAGC -1194
ATCCCTGCAA GGCCTCGGGA GACCCAGAAG TTTCTCGCCC CCTTAGATCC AAACTTGAGC AACCCGGAGT -1124
CTGGATTCCT GGGAAGTCCT CAGCTGTCCT GCGGTTGTGC CGGGGCCCCA GGTCTGGAGG GGACCAGTGG -1054
CCGTGTGGCT TCTACTGCTG GGCTGGAAGT CGGGCCTCCT AGCTCTGCAG TCCGAGGCTT GGAGCCAGGT -984
GCCTGGACCC CGAGGCTGCC CTCCACCCTG TGCGGGCGGG ATGTGACCAG ATGTTGGCCT CATCTGCCAG -914
ACAGAGTGCC GGGGCCCAGG GTCAAGGCCG TTGTGGCTGG TGTGAGGCGC CCGGTGCGCG GCCAGCAGGA -844
CCAC-Box
Sp1
GCGCCTGGCT CCATTTCCCA CCCTTTCTCG ACGGGACCGC CCCGGTGGGT GATTAACAGA TTTGGGGTGG -774
TTTGCTCATG GTGGGGACCC CTCGCCGCCT GAGAACCTGC AAAGAGAAAT GACGGGCCTG TGTCAAGGAG -704
CCCAAGTCGC GGGGAAGTGT TGCAGGGAGG CACTCCGGGA GGTCCCGCGT GCCCGTCCAG GGAGCAATGC -634
AP-2
GTCCTCGGGT TCGTCCCCAG CCGCGTCTAC GCGCCTCCGT CCTCCCCTTC ACGTCCGGCA TTCGTGGTGC -564
CCGGAGCCCG ACGCCCCGCG TCCGGACCTG GAGGCAGCCC TGGGTCTCCG GATCAGGCCA GCGGCCAAAG -494
GGTCGCCGCA CGCACCTGTT CCCAGGGCCT CCACATCATG GCCCCTCCCT CGGGTTACCC CACAGCCTAG -424
Sp1
GCCGATTCGA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCCT GGGAGCGCGA GCGGCGCGCG -354
Sp1
GGCGGGGAAG CGCGGCCCAG ACCCCCGGGT CCGCCCGGAG CAGCTGCGCT GTCGGGGCCA GGCCGGGCTC -284
c-Myc
CCAGTGGATT CGCGGGCACA GACGCCCAGG ACCGCGCTCC CCACGTGGCG GAGGGACTGG GGACCCGGGC -214
Sp1
ACCCGTCCTG CCCCTTCACC TTCCAGCTCC GCCTCCTCCG CGCGGACCCC GCCCCGTCCC GACCCCTCCC -144
Sp1
GGGTCCCCGG CCCAGCCCCC TCCGGGCCCT CCCAGCCCCT CCCCTTCCTT TCCGCGGCCC CGCCCTCTCC -74
c-Myc
TCGCGGCGCG AGTTTCAGGC AGCGCTGCGT CCTGCTGCGC ACGTGGGAAG CCCTGGCCCC GGCCACCCCC -4
GCGATG 3

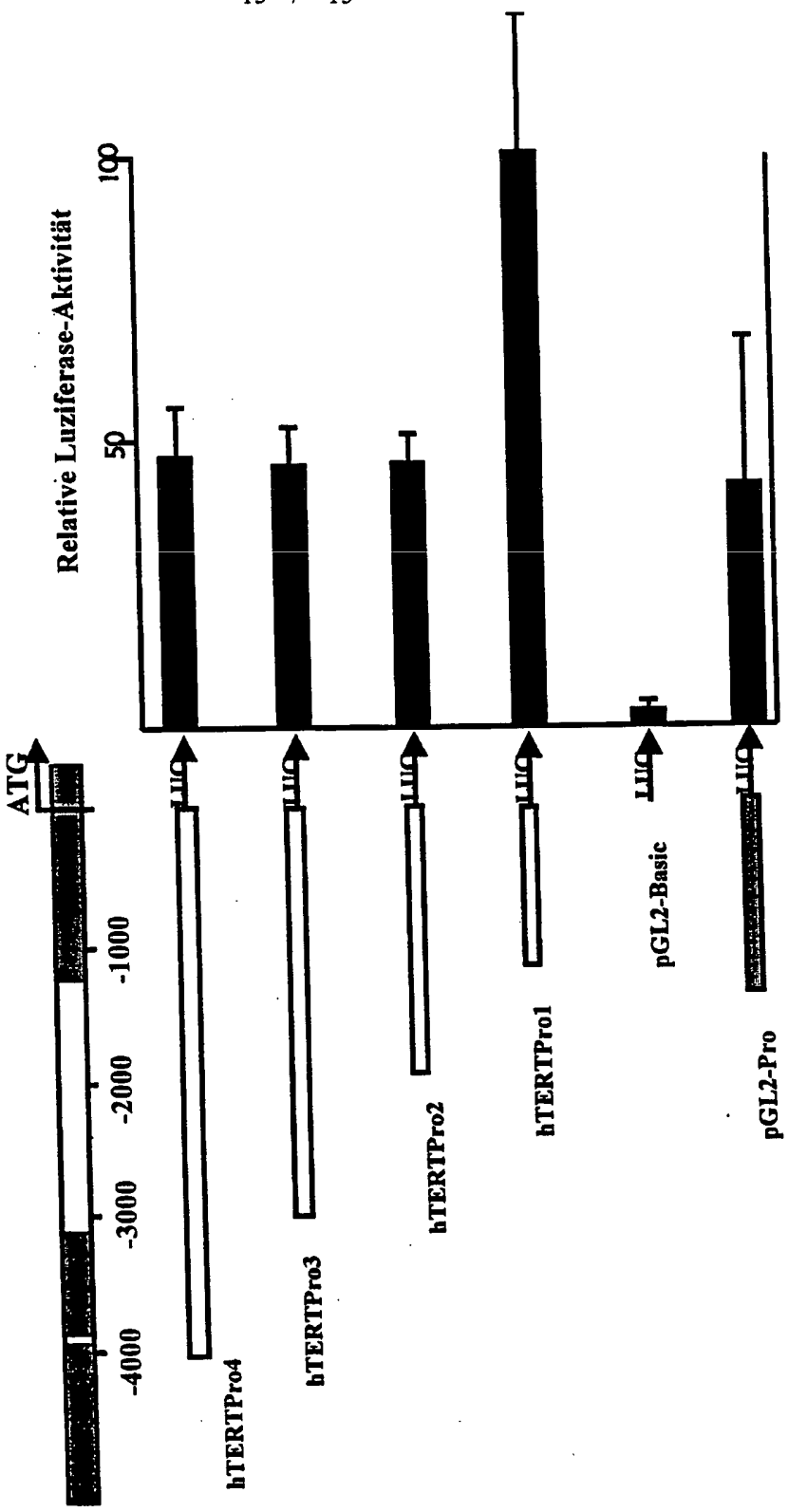

Relative Luziferase-Aktivität
100
50
ATG
-1000
-2000
-3000
-4000
hTERTPro4
hTERTPro3
hTERTPro2
hTERTPro1
pGL2-Basic
pGL2-Pro

Fig.: 11

SEQUENZPROTOKOLL

<110> Bayer AG

<120> Regulatorische DNA-Sequenzen aus der 5i-Region vom Gen der humanen katalytischen Telomerase-Untereinheit und deren diagnostische und therapeutische Verwendung

<130> LeA32805-Ausland

<140>
<141>

<160> 20

<170> PatentIn Vers. 2.0

<210> 1
<211> 5126
<212> DNA
<213> Homo sapiens

<400> 1
gagctctgaa ccgtggaaac gaacatgacc cttgcctgcc tgcttccctg ggtgggtcaa 60
gggtaatgaa gtggtgtgca ggaaatggcc atgtaaatta cacgactctg ctgatgggga 120
ccgttccttc catcattatt catcttcacc cccaaggact gaatgattcc agcaacttct 180
tcgggtgtga caagccatga caaaactcag tacaaacacc actcttttac taggcccaca 240
gagcacgggc cacacccctg atatattaag agtccaggag agatgaggct gctttcagcc 300
accaggctgg ggtgacaaca gcggctgaac agtctgttcc tctagactag tagaccctgg 360
caggcactcc cccaaatttct agggcctggt tgctgcttcc cgagggcgcc atctgccctg 420
gagactcagc ctggggtgcc acactgaggc cagccctgtc tccacaccct ccgcctccag 480
gcctcagctt ctccagcagc ttcctaaacc ctgggtgggc cgtgttccag cgctactgtc 540
tcacctgtcc cactgtgtct tgtctcagcg acgtagctcg cacggttcct cctcacatgg 600
ggtgtctgtc tccttcccca acactcacat gcgttgaagg gaggagattc tgcgcctccc 660
agactggctc ctctgagcct gaacctggct cgtggccccc gatgcaggtt cctggcgtcc 720
ggctgcacgc tgacctccat ttccaggcgc tccccgtctc ctgtcatctg ccggggcctg 780
ccggtgtgtt cttctgtttc tgtgctcctt tccacgtcca gctgcgtgtg tctctgcccg 840
ctagggtctc ggggttttta taggcatagg acgggggcgt ggtgggccag ggcgctcttg 900
ggaaatgcaa catttgggtg tgaaagtagg agtgcctgtc ctcacctagg tccacgggca 960
caggcctggg gatggagccc ccgccaggga cccgcccttc tctgcccagc actttcctgc 1020
ccccctccct ctggaacaca gagtggcagt ttccacaagc actaagcatc ctcttcccaa 1080
aagacccagc attggcaccc ctggacattt gccccacagc cctgggaatt cacgtgacta 1140
cgcacatcat gtacacactc ccgtccacga ccgacccccg ctgttttatt ttaatagcta 1200
caaagcaggg aaatccctgc taaaatgtcc tttaacaaac tggttaaaca aacgggtcca 1260
tccgcacggt ggacagttcc tcacagtgaa gaggaacatg ccgtttataa agcctgcagg 1320
catctcaagg gaattacgct gagtcaaaac tgccacctcc atgggatacg tacgcaacat 1380
gctcaaaaag aaagaatttc accccatggc aggggagtgg ttaggggggt taaggacggt 1440
gggggcggca gctgggggct actgcacgca ccttttacta aagccagttt cctggttctg 1500
atggtattgg ctcagttatg ggagactaac catagggggag tggggatggg ggaacccgga 1560
ggctgtgcca tctttgccat gcccgagtgt cctgggcagg ataatgctct agagatgccc 1620
acgtcctgat tcccccaaac ctgtggacag aacccgcccg gccccagggc ctttgcaggt 1680
gtgatctccg tgaggaccct gaggtctggg atccttcggg actacctgca ggcccgaaaa 1740
gtaatccagg ggttctggga agaggcgggc aggagggtca gaggggggca gcctcaggac 1800
gatggaggca gtcagtctga ggctgaaaag ggagggaggg cctcgagccc aggcctgcaa 1860
gcgcctccag aagctggaaa aagcggggaa gggaccctcc acggagcctg cagcaggaag 1920
gcacggctgg cccttagccc accagggccc atcgtggacc tccggcctcc gtgccatagg 1980
agggcactcg cgctgccctt ctagcatgaa gtgtgtgggg atttgcagaa gcaacaggaa 2040
acccatgcac tgtgaatcta ggattatttc aaaacaaagg tttacagaaa catccaagga 2100
cagggctgaa gtgcctccgg gcaagggcag ggcaggcacg agtgatttta tttagctatt 2160
ttattttatt tacttacttt ctgagacaga gttatgctct tgttgcccag gctggagtgc 2220
agcggcatga tcttggctca ctgcaacctc cgtctcctgg gttcaagcaa ttctcgtgcc 2280
tcagcctccc aagtagctgg gatttcaggc gtgcaccacc acacccggct aattttgtat 2340
ttttagtaga gatgggcttt caccatgttg gtcaagctga tctcaaaatc ctgacctcag 2400
gtgatccgcc cacctcagcc tcccaaagtg ctgggattac aggcatgagc cactgcacct 2460

ggcctattta accattttaa aacttccctg ggctcaagtc acacccactg gtaaggagtt 2520
catggagttc aatttcccct ttactcagga gttaccctcc tttgatattt tctgtaattc 2580
ttcgtagact ggggatacac cgtctcttga catattcaca gtttctgtga ccacctgtta 2640
tcccatggga cccactgcag gggcagctgg gaggctgcag gcttcaggtc ccagtggggt 2700
tgccatctgc cagtagaaac ctgatgtaga atcagggcgc aagtgtggac actgtcctga 2760
atctcaatgt ctcagtgtgt gctgaaacat gtagaaatta aagtccatcc ctcctactct 2820
actgggattg agccccttcc ctatcccccc ccaggggcag aggagttcct ctcactcctg 2880
tggaggaagg aatgatactt tgttattttt cactgctggt actgaatcca ctgtttcatt 2940
tgttggtttg tttgttttgt tttgagaggc ggtttcactc ttgttgctca ggctggaggg 3000
agtgcaatgg cgcgatcttg gcttactgca gcctctgcct cccaggttca agtgattctc 3060
ctgcttccgc ctcccatttg gctgggatta caggcacccg ccaccatgcc cagctaattt 3120
tttgtatttt tagtagagac gggggtgggt ggggttcacc atgttggcca ggctggtctc 3180
gaacttctga cctcagatga tccacctgcc tctgcctcct aaagtgctgg gattacaggt 3240
gtgagccacc atgcccagct cagaatttac tctgtttaga aacatctggg tctgaggtag 3300
gaagctcacc ccactcaagt gttgtggtgt tttaagccaa tgatagaatt tttttattgt 3360
tgttagaaca ctcttgatgt tttacactgt gatgactaag acatcatcag ctttttcaaag 3420
acacactaac tgcacccata atactggggt gtcttctggg tatcagcaat cttcattgaa 3480
tgccgggagg cgtttcctcg ccatgcacat ggtgttaatt actccagcat aatcttctgc 3540
ttccatttct tctcttccct cttttaaaat tgtgttttct atgttggctt ctctgcagag 3600
aaccagtgta agctacaact taacttttgt tggaacaaat tttccaaacc gccccctttgc 3660
cctagtggca gagacaattc acaaacacag ccctttaaaa aggcttaggg atcactaagg 3720
ggatttctag aagagcgacc tgtaatccta agtatttaca agacgaggct aacctccagc 3780
gagcgtgaca gcccagggag ggtgcgaggc ctgttcaaat gctagctcca taaataaagc 3840
aatttcctcc ggcagtttct gaaagtagga aaggttacat ttaaggttgc gtttgttagc 3900
atttcagtgt ttgccgacct cagctacagc atccctgcaa ggcctcggga gacccagaag 3960
ttctctcgcc ccttagatcc aaacttgagc aacccggagt ctggattcct gggaagtcct 4020
cagctgtcct gcggttgtgc cggggcccca ggtctggagg ggaccagtgg ccgtgtggct 4080
tctactgctg ggctggaagt cgggcctcct agctctgcag tccgaggctt ggagccaggt 4140
gcctggaccc cgaggctgcc ctccaccctg tgcgggcggg atgtgaccag atgttggcct 4200
catctgccag acagagtgcc ggggcccagg gtcaaggccg ttgtggctgg tgtgaggcgc 4260
ccggtgcgcg gccagcagga gcgcctggct ccatttccca ccctttctcg acgggaccgc 4320
cccggtgggt gattaacaga tttggggtgg tttgctcatg gtggggaccc ctcgccgcct 4380
gagaacctgc aaagagaaat gacgggcctg tgtcaaggag cccaagtcgc ggggaagtgt 4440
tgcagggagg cactccggga ggtcccgcgt gcccgtccag ggagcaatgc gtcctcgggt 4500
tcgtccccag ccgcgtctac gcgcctccgt cctccccttc acgtccggca ttcgtggtgc 4560
ccggagcccg acgccccgcg tccggacctg gaggcagccc tgggtctccg gatcaggcca 4620
gcggccaaag ggtcgccgca cgcacctgtt cccagggcct ccacatcatg gcccctccct 4680
cgggttaccc cacagcctag gccgattcga cctctctccg ctggggccct cgctggcgtc 4740
cctgcaccct gggagcgcga gcggcgcgcg ggcggggaag cgcggcccag accccgggt 4800
ccgcccggag cagctgcgct gtcggggcca ggccgggctc ccagtggatt cgcgggcaca 4860
gacgcccagg accgcgctcc ccacgtggcg gagggactgg ggacccgggc acccgtcctg 4920
ccccttcacc ttccagctcc gcctcctccg cgcggacccc gccccgtccc gacccctccc 4980
gggtccccgg cccagccccc tccgggccct cccagccccc ccccttcctt tccgcggccc 5040
cgccctctcc tcgcggcgcg agtttcaggc agcgctgcgt cctgctgcgc acgtgggaag 5100
ccctggcccc ggccaccccc gcgatg 5126

<210> 2
<211> 4042
<212> DNA
<213> Homo sapiens

<400> 2
gtttcaggca gcgctgcgtc ctgctgcgca cgtgggaagc cctggccccg gccacccccg 60
cgatgccgcg cgctccccgc tgccgagccg tgcgctccct gctgcgcagc cactaccgcg 120
aggtgctgcc gctggccacg ttcgtgcggc gcctggggcc ccagggctgg cggctggtgc 180
agcgcgggga cccggcggct ttccgcgcgc tggtggccca gtgcctggtg tgcgtgccct 240
gggacgcacg gccgcccccc gccgcccccc ccttccgcca ggtgtcctgc ctgaaggagc 300
tggtggcccg agtgctgcag aggctgtgcg agcgcggcgc gaagaacgtg ctggccttcg 360
gcttcgcgct gctggacggg gcccgcgggg gccccccccga ggccttcacc accagcgtgc 420
gcagctacct gcccaacacg gtgaccgacg cactgcgggg gagcggggcg tgggggctgc 480
tgctgcgccg cgtgggcgac gacgtgctgg ttcacctgct ggcacgctgc gcgctctttg 540
tgctggtggc tcccagctgc gcctaccagg tgtgcgggcc gccgctgtac cagctcggcg 600
ctgccactca ggcccggccc ccgccacacg ctagtggacc ccgaaggcgt ctgggatgcg 660
aacgggcctg gaaccatagc gtcagggagg ccggggtccc cctgggcctg ccagccccgg 720
gtgcgaggag gcgcgggggc agtgccagcc gaagtctgcc gttgcccaag aggcccaggc 780

gtggcgctgc ccctgagccg gagcggacgc ccgttgggca ggggtcctgg gcccacccgg 840
gcaggacgcg tggaccgagt gaccgtggtt tctgtgtggt gtcacctgcc agacccgccg 900
aagaagccac ctctttggag ggtgcgctct ctggcacgcg ccactcccac ccatccgtgg 960
gccgccagca ccacgcgggc cccccatcca catcgcggcc accacgtccc tgggacacgc 1020
cttgtccccc ggtgtacgcc gagaccaagc acttcctcta ctcctcaggc gacaaggagc 1080
agctgcggcc ctccttcctа ctcagctctc tgaggcccag cctgactggc gctcggaggc 1140
tcgtggagac catctttctg ggttccaggc cctggatgcc agggactccc cgcaggttgc 1200
cccgcctgcc ccagcgctac tggcaaatgc ggcccctgtt tctggagctg cttgggaacc 1260
acgcgcagtg cccctacggg gtgctcctca agacgcactg cccgctgcga gctgcggtca 1320
ccccagcagc cggtgtctgt gcccgggaga agccccaggg ctctgtggcg gcccccgagg 1380
aggaggacac agacccccgt cgcctggtgc agctgctccg ccagcacagc agcccctggc 1440
aggtgtacgg cttcgtgcgg gcctgcctgc gccggctggt gcccccaggc ctctggggct 1500
ccaggcacaa cgaacgccgc ttcctcagga acaccaagaa gttcatctcc ctggggaagc 1560
atgccaagct ctcgctgcag gagctgacgt ggaagatgag cgtgcgggac tgcgcttggc 1620
tgcgcaggag cccaggggtt ggctgtgttc cggccgcaga gcaccgtctg cgtgaggaga 1680
tcctggccaa gttcctgcac tggctgatga gtgtgtacgt cgtcgagctg ctcaggtctt 1740
tcttttatgt cacggagacc acgtttcaaa agaacaggct ctttttctac cggaagagtg 1800
tctggagcaa gttgcaaagc attggaatca gacagcactt gaagagggtg cagctgcggg 1860
agctgtcgga agcagaggtc aggcagcatc gggaagccag gcccgccctg ctgacgtcca 1920
gactccgctt catccccaag cctgacgggc tgcggccgat tgtgaacatg gactacgtcg 1980
tgggagccag aacgttccgc agagaaaaga gggccgagcg tctcacctcg agggtgaagg 2040
cactgttcag cgtgctcaac tacgagcggg cgcggcgccc cggcctcctg ggcgcctctg 2100
tgctgggcct ggacgatatc cacagggcct ggcgcacctt cgtgctgcgt gtgcgggccc 2160
aggacccgcc gcctgagctg tactttgtca aggtggatgt gacgggcgcg tacgacacca 2220
tccccagga caggctcacg gaggtcatcg ccagcatcat caaaccccag aacacgtact 2280
gcgtgcgtcg gtatgccgtg gtccagaagg ccgcccatgg gcacgtccgc aaggccttca 2340
agagccacgt ctctaccttg acagacctcc agccgtacat gcgacagttc gtggctcacc 2400
tgcaggagac cagcccgctg agggatgccg tcgtcatcga gcagagctcc tccctgaatg 2460
aggccagcag tggcctcttc gacgtcttcc tacgcttcat gtgccaccac gccgtgcgca 2520
tcaggggcaa gtcctacgtc cagtgccagg ggatcccgca gggctccatc ctctccacgc 2580
tgctctgcag cctgtgctac ggcgacatgg agaacaagct gtttgcgggg attcggcggg 2640
acgggctgct cctgcgtttg gtggatgatt tcttgttggt gacacctcac ctcacccacg 2700
cgaaaacctt cctcaggacc ctggtccgag gtgtccctga gtatggctgc gtggtgaact 2760
tgcggaagac agtggtgaac ttccctgtag aagacgaggc cctgggtggc acggcttttg 2820
ttcagatgcc ggcccacggc ctattcccct ggtgcggcct gctgctggat acccggaccc 2880
tggaggtgca gagcgactac tccagctatg cccggacctc catcagagcc agtctcacct 2940
tcaaccgcgg cttcaaggct gggaggaaca tgcgtcgcaa actctttggg gtcttgcggc 3000
tgaagtgtca cagcctgttt ctggatttgc aggtgaacag cctccagacg gtgtgcacca 3060
acatctacaa gatcctcctg ctgcaggcgt acaggtttca cgcatgtgtg ctgcagctcc 3120
catttcatca gcaagtttgg aagaacccca cattttcct gcgcgtcatc tctgacacgg 3180
cctccctctg ctactccatc ctgaaagcca agaacgcagg gatgtcgctg ggggccaagg 3240
gcgccgccgg ccctctgccc tccgaggccg tgcagtggct gtgccaccaa gcattcctgc 3300
tcaagctgac tcgacaccgt gtcacctacg tgccactcct ggggtcactc aggacagccc 3360
agacgcagct gagtcggaag ctcccgggga cgacgctgac tgccctggag gccgcagcca 3420
acccggcact gccctcagac ttcaagacca tcctggactg atggccaccc gcccacagcc 3480
aggccgagag cagacaccag cagccctgtc acgccgggct ctacgtccca gggagggagg 3540
ggcggcccac acccaggccc gcaccgctgg gagtctgagg cctgagtgag tgtttggccg 3600
aggcctgcat gtccggctga aggctgagtg tccggctgag gcctgagcga gtgtccagcc 3660
aagggctgag tgtccagcac acctgccgtc ttcacttccc cacaggctgg cgctcggctc 3720
caccccaggg ccagcttttc ctcaccagga gcccggcttc cactccccac ataggaatag 3780
tccatcccca gattcgccat tgttcacccc tcgccctgcc ctcctttgcc ttccaccccc 3840
accatccagg tggagaccct gagaaggacc ctgggagctc tgggaatttg gagtgaccaa 3900
aggtgtgccc tgtacacagg cgaggaccct gcacctggat gggggtccct gtgggtcaaa 3960
ttgggggggag gtgctgtggg agtaaaatac tgaatatatg agtttttcag ttttgaaaaa 4020
aaaaaaaaaa aaaaaaaaaa aa 4042

<210> 3
<211> 11276
<212> DNA
<213> Homo sapiens

<400> 3
acttgagccc aagagttcaa ggctacggtg agccatgatt gcaacaccac acgccagcct 60
tggtgacaga atgagaccct gtctcaaaaa aaaaaaaaaa aattgaaata atataaagca 120
tcttctctgg ccacagtgga acaaaaccag aaatcaacaa caagaggaat tttgaaaact 180

atacaaacac atgaaaatta aacaatatac ttctgaatga ccagtgagtc aatgaagaaa 240
ttaaaaagga aattgaaaaa tttatttaag caaatgataa cggaaacata acctctcaaa 300
acccacggta tacagcaaaa gcagtgctaa gaaggaagtt tatagctata agcagctaca 360
tcaaaaaagt agaaaagcca ggcgcagtgg ctcatgcctg taatcccagc actttgggag 420
gccaaggcgg gcagatcgcc tgaggtcagg agttcgagac cagcctgacc aacacagaga 480
aaccttgtcg ctactaaaaa tacaaaatta gctgggcatg gtggcacatg cctgtaatcc 540
cagctactcg ggaggctgag gcaggataac cgcttgaacc caggaggtgg aggttgcggt 600
gagccgggat tgcgccattg gactccagcc tgggtaacaa gagtgaaacc ctgtctcaag 660
aaaaaaaaaa aagtagaaaa acttaaaaat acaacctaat gatgcacctt aaagaactag 720
aaaagcaaga gcaaactaaa cctaaaattg gtaaaagaaa agaaataata aagatcagag 780
cagaaataaa tgaaactgaa agataacaat acaaaagatc aacaaaatta aaagttggtt 840
ttttgaaaag ataaacaaaa ttgacaaacc tttgcccaga ctaagaaaaa aggaaagaag 900
acctaaataa ataaagtcag agatgaaaaa agagacatta caactgatac cacagaaatt 960
caaaggatca ctagaggcta ctatgagcaa ctgtacacta ataaattgaa aaacctagaa 1020
aaaatagata aattcctaga tgcatacaac ctaccaagat tgaaccatga agaaatccaa 1080
agcccaaaca gaccaataac aataatggga ttaaagccat aataaaaagt ctcctagcaa 1140
agagaagccc aggacccaat ggcttccctg ctggatttta ccaatcattt aaagaagaat 1200
gaattccaat cctactcaaa ctattctgaa aaatagagga aagaatactt ccaaactcat 1260
tctacatggc cagtattacc ctgattccaa aaccagacaa aaacacatca aaaacaaaca 1320
aacaaaaaaa cagaaagaaa gaaaactaca ggccaatatc cctgatgaat actgatacaa 1380
aaatcctcaa caaaacacta gcaaaccaaa ttaaacaaca ccttcgaaag atcattcatt 1440
gtgatcaagt gggatttatt ccagggatgg aaggatggtt caacatatgc aaatcaatca 1500
atgtgataca tcatcccaac aaaatgaagt acaaaaacta tatgattatt tcactttatg 1560
cagaaaaagc atttgataaa attctgcacc cttcatgata aaaaccctca aaaaaccagg 1620
tatacaagaa acatacaggc caggcacagt ggctcacacc tgcgatccca gcactctggg 1680
aggccaaggt gggatgattg cttgggccca ggagtttgag actagcctgg gcaacaaaat 1740
gagacctggt ctacaaaaaa ctttttaaa aaattagcca ggcatgatgg catatgcctg 1800
tagtcccagc tagtctggag gctgaggtgg gagaatcact taagcctagg aggtcgaggc 1860
tgcagtgagc catgaacatg tcactgtact ccagcctaga caacagaaca agaccccact 1920
gaataagaag aaggagaagg agaagggaga agggagggag aagggaggag gaggagaagg 1980
aggaggtgga ggagaagtgg aaggggaagg ggaagggaaa gaggaagaag aagaaacata 2040
tttcaacata ataaaagccc tatatgacag accgaggtag tattatgagg aaaaactgaa 2100
agcctttcct ctaagatctg gaaaatgaca agggcccact ttcaccactg tgattcaaca 2160
tagtactaga agtcctagct agagcaatca gataagagaa agaaataaaa ggcatccaaa 2220
ctggaaagga agaagtcaaa ttatcctgtt tgcagatgat atgatcttat atctggaaaa 2280
gacttaagac accactaaaa aactattaga gctgaaattt ggtacagcag gatacaaaat 2340
caatgtacaa aaatcagtag tatttctata ttccaacagc aaacaatctg aaaaagaaac 2400
caaaaaagca gctacaaata aaattaaaca gctaggaatt aaccaaagaa gtgaaagatc 2460
tctacaatga aaactataaa atgttgataa aagaaattga agagggcaca aaaaaagaaa 2520
agatattcca tgttcataga ttggaagaat aaatactgtt aaaatgtcca tactacccaa 2580
agcaatttac aaattcaatg caatccctat taaaatacta atgacgttct tcacagaaat 2640
agaagaaaca attctaagat ttgtacagaa ccacaaaaga cccagaatag ccaaagctat 2700
cctgaccaaa aagaacaaaa ctggaagcat cacattacct gacttcaaat tatactacaa 2760
agctatagta acccaaacta catggtactg gcataaaaac agatgagaca tggaccagag 2820
gaacagaata gagaatccag aaacaaatcc atgcatctac agtgaactca tttttgacaa 2880
aggtgccaag aacatacttt ggggaaaaga taatctcttc aataaatggt gctggaggaa 2940
ctggatatcc atatgcaaaa taacaatact agaactctgt ctctcaccat atacaaaagc 3000
aaatcaaaat ggatgaaagg cttaaatcta aaacctcaaa ctttgcaact actaaaagaa 3060
aacaccggag aaactctcca ggacattgga gtgggcaaag actttcttgag taattccctg 3120
caggcacagg caaccaaagc aaaaacagac aaatgggatc atatcaagtt aaaaagcttc 3180
tgcccagcaa aggaaacaat caacaaagag aagagacaac ccacagaatg ggagaatata 3240
tttgcaaact attcatctaa caaggaatta ataaccagta tatataagga gctcaaacta 3300
ctctataaga aaaacaccta ataagctgat tttcaaaaat aagcaaaaga tctgggtaga 3360
catttctcaa aataagtcat acaaatggca aacaggcatc tgaaaatgtg ctcaacacca 3420
ctgatcatca gagaaatgca aatcaaaact actatgagag atcatctcat cccagttaaa 3480
atggctttta ttcaaaagac aggcaataac aaatgccagt gaggatgtgg ataaaaggaa 3540
acccttggac actgttggtg ggaatggaaa ttgctaccac tatggagaac agtttgaaag 3600
ttcctcaaaa aactaaaaat aaagctacca tacagcaatc ccattgctag gtatatactc 3660
caaaaaaggg aatcagtgta tcaacaagct atctccactc ccacatttac tgcagcactg 3720
ttcatagcag ccaaggtttg gaagcaacct cagtgtccat caacagacga atggaaaaag 3780
aaaatgtggt gcacatacac aatggagtac tacgcagcca taaaaaagaa tgagatcctg 3840
tcagttgcaa cagcatgggg ggcactggtc agtatgttaa gtgaaataag ccaggcacag 3900
aaagacaaac ttttcatgtt ctcccttact tgtgggagca aaaattaaaa caattgacat 3960
agaaatagag gagaatggtg gttctagagg ggtgggggac agggtgacta gagtcaacaa 4020
taatttattg tatgttttaa aataactaaa agagtataat tgggttgttt gtaacacaaa 4080

gaaaggataa atgcttgaag gtgacagata ccccatttac cctgatgtga ttattacaca 4140
ttgtatgcct gtatcaaaat atctcatgta tgctatagat ataaacccta ctatattaaa 4200
aattaaaatt ttaatggcca ggcacggtgg ctcatgtccg taatcccagc actttgggag 4260
gccgaggcgg gtggatcacc tgaggtcagg agtttgaaac cagtctggcc accatgatga 4320
aaccctgtct ctactaaaga tacaaaaatt agccaggcgt ggtggcacat acctgtagtc 4380
ccaactactc aggaggctga gacaggagaa ttgcttgaac ctgggaggcg gaggttgcag 4440
tgagccgaga tcatgccact gcactgcagc ctgggtgaca gagcaagact ccatctcaaa 4500
acaaaaacaa aaaaagaag attaaaattg taatttttat gtaccgtata aatatatact 4560
ctactatatt agaagttaaa aattaaaaca attataaaag gtaattaacc acttaatcta 4620
aaataagaac aatgtatgtg gggtttctag cttctgaaga agtaaaagtt atggccacga 4680
tggcagaaat gtgaggaggg aacagtggaa gttactgttg ttagacgctc atactctctg 4740
taagtgactt aattttaacc aaagacaggc tgggagaagt taaagaggca ttctataagc 4800
cctaaaacaa ctgctaataa tggtgaaagg taatctctat taattaccaa taattacaga 4860
tatctctaaa atcgagctgc agaattggca cgtctgatca caccgtcctc tcattcacgg 4920
tgctttttt cttgtgtgct tggagatttt cgattgtgtg ttcgtgtttg gttaaactta 4980
atctgtatga atcctgaaac gaaaaatggt ggtgatttcc tccagaagaa ttagagtacc 5040
tggcaggaag caggtggctc tgtggacctg agccacttca atcttcaagg gtctctggcc 5100
aagacccagg tgcaaggcag aggcctgatg acccgaggac aggaaagctc ggatgggaag 5160
gggcgatgag aagcctgcct cgttggtgag cagcgcatga agtgccctta tttacgcttt 5220
gcaaagattg ctctggatac catctggaaa aggcggccag cgggaatgca aggagtcaga 5280
agcctcctgc tcaaacccag gccagcagct atggcgccca cccgggcgtg tgccagaggg 5340
agaggagtca aggcacctcg aagtatggct taaatctttt tttcacctga agcagtgacc 5400
aaggtgtatt ctgagggaag cttgagttag gtgccttctt taaaacagaa agtcatggaa 5460
gcaccctct caagggaaaa ccagacgccc gctctgcggt catttacctc tttcctctct 5520
ccctctcttg ccctcgcggt ttctgatcgg gacagagtga cccccgtgga gcttctccga 5580
gcccgtgctg aggaccctct tgcaaagggc tccacagacc cccgccctgg agagaggagt 5640
ctgagcctgg cttaataaca aactgggatg tggctggggg cggacagcga cggcgggatt 5700
caaagactta attccatgag taaattcaac ctttccacat ccgaatggat ttggatttta 5760
tcttaatatt ttcttaaatt tcatcaaata acattcagga ctgcagaaat ccaaaggcgt 5820
aaaacaggaa ctgagctatg tttgccaagg tccaaggact taataaccat gttcagaggg 5880
atttttcgcc ctaagtactt tttattggtt ttcataaggt ggcttagggt gcaagggaaa 5940
gtacacgagg agaggcctgg gcggcagggc tatgagcacg gcagggccac cggggagaga 6000
gtccccggcc tgggaggctg acagcaggac cactgaccgt cctccctggg agctgccaca 6060
ttgggcaacg cgaaggcggc cacgctgcgt gtgactcagg accccatacc ggcttcctgg 6120
gcccacccac actaacccag gaagtcacgg agctctgaac ccgtggaaac gaacatgacc 6180
cttgcctgcc tgcttccctg ggtgggtcaa gggtaatgaa gtggtgtgca ggaaatggcc 6240
atgtaaatta cacgactctg ctgatgggga ccgttccttc catcattatt catcttcacc 6300
cccaaggact gaatgattcc agcaacttct tcgggtgtga caagccatga caaaactcag 6360
tacaaacacc actcttttac taggcccaca gagcacggsc cacacccctg atatattaag 6420
agtccaggag agatgaggct gctttcagcc accaggctgg ggtgacaaca gcggctgaac 6480
agtctgttcc tctagactag tagaccctgg caggcactcc cccagattct agggcctggt 6540
tgctgcttcc cgagggcgcc atctgccctg gagactcagc ctggggtgcc acactgaggc 6600
cagccctgtc tccacaccct ccgcctccag gcctcagctt ctccagcagc ttcctaaacc 6660
ctgggtgggc cgtgttccag cgctactgtc tcacctgtcc cactgtgtct tgtctcagcg 6720
acgtagctcg cacggttcct cctcacatgg ggtgtctgtc tccttcccca acactcacat 6780
gcgttgaagg gaggagattc tgcgcctccc agactggctc ctctgagcct gaacctggct 6840
cgtggccccc gatgcaggtt cctggcgtcc ggctgcacgc tgacctccat ttccaggcgc 6900
tccccgtctc ctgtcatctg ccggggcctg ccggtgtgtt cttctgtttc tgtgctcctt 6960
tccacgtcca gctgcgtgtg tctctgcccg ctagggtctc ggggttttta taggcatagg 7020
acgggggcgt ggtgggccag ggcgctcttg ggaaatgcaa catttgggtg tgaaagtagg 7080
agtgcctgtc ctcacctagg tccacgggca caggcctggg gatggagccc ccgccaggga 7140
cccgcccttc tctgcccagc actttcctgc ccccctccct ctggaacaca gagtggcagt 7200
ttccacaagc actaagcatc ctcttcccaa aagacccagc attggcaccc ctggacattt 7260
gccccacagc cctgggaatt cacgtgacta cgcacatcat gtacacactc ccgtccacga 7320
ccgaccccg ctgttttatt ttaatagcta caaagcaggg aaatccctgc taaaatgtcc 7380
tttaacaaac tggttaaaca aacgggtcca tccgcacggt ggacagttcc tcacagtgaa 7440
gaggaacatg ccgtttataa agcctgcagg catctcaagg gaattacgct gagtcaaaac 7500
tgccacctcc atgggatacg tacgcaacat gctcaaaaag aaagaatttc accccatggc 7560
aggggagtgg ttaggggggt taaggacggt gggggcggca gctgggggct actgcacgca 7620
cctttacta aagccagttt cctggttctg atggtattgg ctcagttatg ggagactaac 7680
catagggag tggggatggg ggaacccgga ggctgtgcca tctttgccat gcccgagtgt 7740
cctgggcagg ataatgctct agagatgccc acgtcctgat tcccccaaac ctgtggacag 7800
aacccgcccg gccccagggc ctttgcaggt gtgatctccg tgaggaccct gaggtctggg 7860
atccttcggg actacctgca ggcccgaaaa gtaatccagg ggttctggga agaggcgggc 7920
aggagggtca gaggggggca gcctcaggac gatggaggca gtcagtctga ggctgaaaag 7980

ggagggaggg cctcgagccc aggcctgcaa gcgcctccag aagctggaaa aagcggggaa 8040
gggaccctcc acggagcctg cagcaggaag gcacggctgg cccttagccc accagggccc 8100
atcgtggacc tccggcctcc gtgccatagg agggcactcg cgctgccctt ctagcatgaa 8160
gtgtgtgggg atttgcagaa gcaacaggaa acccatgcac tgtgaatcta ggattatttc 8220
aaaacaaagg tttacagaaa catccaagga cagggctgaa gtgcctccgg gcaagggcag 8280
ggcaggcacg agtgatttta tttagctatt ttatttttatt tacttacttt ctgagacaga 8340
gttatgctct tgttgcccag gctggagtgc agcggcatga tcttggctca ctgcaacctc 8400
cgtctcctgg gttcaagcaa ttctcgtgcc tcagcctccc aagtagctgg gatttcaggc 8460
gtgcaccacc acacccggct aattttgtat ttttagtaga gatgggcttt caccatgttg 8520
gtcaagctga tctcaaaatc ctgacctcag gtgatccgcc cacctcagcc tcccaaagtg 8580
ctgggattac aggcatgagc cactgcacct ggcctattta accattttaa aacttccctg 8640
ggctcaagtc acacccactg gtaaggagtt catggagttc aatttccccct ttactcagga 8700
gttaccctcc tttgatattt tctgtaattc ttcgtagact ggggatacac cgtctcttga 8760
catattcaca gtttctgtga ccacctgtta tcccatggga cccactgcag gggcagctgg 8820
gaggctgcag gcttcaggtc ccagtggggt tgccatctgc cagtagaaac ctgatgtaga 8880
atcagggcgc aagtgtggac actgtcctga atctcaatgt ctcagtgtgt gctgaaacat 8940
gtagaaatta aagtccatcc ctcctactct actgggattg agccccttcc ctatcccccc 9000
ccaggggcag aggagttcct ctcactcctg tggaggaagg aatgatactt tgttattttt 9060
cactgctggt actgaatcca ctgtttcatt tgttggtttg tttgttttgt tttgagaggc 9120
ggtttcactc ttgttgctca ggctggaggg agtgcaatgg cgcgatcttg gcttactgca 9180
gcctctgcct cccaggttca agtgattctc ctgcttccgc ctcccatttg gctgggatta 9240
caggcacccg ccaccatgcc cagctaattt tttgtatttt tagtagagac gggggtgggt 9300
ggggttcacc atgttggcca ggctggtctc gaacttctga cctcagatga tccacctgcc 9360
tctgcctcct aaagtgctgg gattacaggt gtgagccacc atgcccagct cagaatttac 9420
tctgtttaga aacatctggg tctgaggtag gaagctcacc ccactcaagt gttgtggtgt 9480
tttaagccaa tgatagaatt tttttattgt tgttagaaca ctcttgatgt tttacactgt 9540
gatgactaag acatcatcag cttttcaaag acacactaac tgcacccata atactggggt 9600
gtcttctggg tatcagcaat cttcattgaa tgccgggagg cgtttcctcg ccatgcacat 9660
ggtgttaatt actccagcat aatcttctgc ttccatttct tctcttccct cttttaaaat 9720
tgtgttttct atgttggctt ctctgcagag aaccagtgta agctacaact taacttttgt 9780
tggaacaaat tttccaaacc gccccttgc cctagtggca gagacaattc acaaacacag 9840
ccctttaaaa aggcttaggg atcactaagg ggatttctag aagagcgacc tgtaatccta 9900
agtatttaca agacgaggct aacctccagc gagcgtgaca gcccagggag ggtgcgaggc 9960
ctgttcaaat gctagctcca taaataaagc aatttcctcc ggcagtttct gaaagtagga 10020
aaggttacat ttaaggttgc gtttgttagc atttcagtgt ttgccgacct cagctacagc 10080
atccctgcaa ggcctcggga gacccagaag tttctcgccc ccttagatcc aaacttgagc 10140
aacccggagt ctggattcct gggaagtcct cagctgtcct gcggttgtgc cggggcccca 10200
ggtctggagg ggaccagtgg ccgtgtggct tctactgctg ggctggaagt cgggcctcct 10260
agctctgcag tccgaggctt ggagccaggt gcctggaccc cgaggctgcc ctccaccctg 10320
tgcgggcggg atgtgaccag atgttggcct catctgccag acagagtgcc ggggcccagg 10380
gtcaaggccg ttgtggctgg tgtgaggcgc ccggtgcgcg gccagcagga gcgcctggct 10440
ccatttccca ccctttctcg acgggaccgc cccggtgggt gattaacaga tttggggtgg 10500
tttgctcatg gtggggaccc ctcgccgcct gagaacctgc aaagagaaat gacgggcctg 10560
tgtcaaggag cccaagtcgc ggggaagtgt tgcagggagg cactccggga ggtcccgcgt 10620
gcccgtccag ggagcaatgc gtcctcgggt tcgtccccag ccgcgtctac gcgcctccgt 10680
cctccccttc acgtccggca ttcgtggtgc ccggagcccg acgccccgcg tccggacctg 10740
gaggcagccc tgggtctccg gatcaggcca gcggccaaag ggtcgccgca cgcacctgtt 10800
cccagggcct ccacatcatg gcccctcccc cgggttaccc cacagcctag gccgattcga 10860
cctctctccg ctggggccct cgctggcgtc cctgcaccct gggagcgcga gcggcgcgcg 10920
ggcggggaag cgcggcccag acccccgggt ccgcccggag cagctgcgct gtcggggcca 10980
ggccgggctc ccagtggatt cgcgggcaca gacgcccagg accgcgctcc ccacgtggcg 11040
gagggactgg ggacccgggc acccgtcctg ccccttcacc ttccagctcc gcctcctccg 11100
cgcggacccc gccccgtccc gacccctccc gggtccccgg cccagccccc tccgggccct 11160
cccagcccct ccccttcctt tccgcggccc cgccctctcc tcgcggcgcg agtttcaggc 11220
agcgctgcgt cctgctgcgc acgtgggaag ccctggcccc ggccaccccc gcgatg 11276

<210> 4
<211> 104
<212> DNA
<213> Homo sapiens

<400> 4
gtgggcctcc ccggggtcgg cgtccggctg gggttgaggg cggccggggg gaaccagcga 60
catgcggaga gcagcgcagg cgactcaggg cgcttccccc gcag 104

<210> 5
<211> 8616
<212> DNA
<213> Homo sapiens

<400> 5
gtgaggaggt ggtggccgtc gagggcccag gccccagagc tgaatgcagt aggggctcag 60
aaaagggggc aggcagagcc ctggtcctcc tgtctccatc gtcacgtggg cacacgtggc 120
ttttcgctca ggacgtcgag tggacacggt gatctctgcc tctgctctcc ctcctgtcca 180
gtttgcataa acttacgagg ttcaccttca cgttttgatg gacacgcggt ttccaggcgc 240
cgaggccaga gcagtgaaca gaggaggctg ggcgcggcag tggagccggg ttgccggcaa 300
tggggagaag tgtctggaag cacagacgct ctggcgaggg tgcctgcagg ttacctataa 360
tcctcttcgc aatttcaagg gtgggaatga gaggtgggga cgagaacccc ctcttcctgg 420
gggtgggagg taagggtttt gcaggtgcac gtggtcagcc aatatgcagg tttgtgttta 480
agatttaatt gtgtgttgac ggccaggtgc ggtggctcac gccggtaatc ccagcacttt 540
gggaagctga ggcaggtgga tcacctgagg tcaggagttt gagaccagcc tgaccaacat 600
ggtgaaaccc tatctgtact aaaaatacaa aaattagctg ggcatggtgg tgtgtgcctg 660
taatcccagc tacttgggag gctgaggcag gagaatcact tgaacccagg aggcggaggc 720
tgcagtgagc tgagattgtg ccattgtact ccagcctggg cgacaagagt gaaactctgt 780
ctttaaaaaa aaaaagtgtt cgttgattgt gccaggacag ggtagaggga gggagataag 840
actgttctcc agcacagatc ctggtcccat ctttaggtat gaagagggcc acatgggagc 900
agaggacagc agatggctcc acctgctgag gaagggacag tgtttgtggg tgttcagggg 960
atggtgctgc tgggccctgc cgtgtcccca ccctgttttt ctggatttga tgttgaggaa 1020
cctccgctcc agcccccttt tggctcccag tgctcccagg ccctaccgtg gcagctagaa 1080
gaagtcccga tttcaccccc tccccacaaa ctcccaagac atgtaagact tccggccatg 1140
cagacaagga gggtgacctt cttggggctc ttttttttct ttttttcttt ttatggtggc 1200
aaaagtcata taacatgaga ttggcactcc taacaccgtt ttctgtgtac agtgcagaat 1260
tgctaactcg gcggtgttta cagcaggttg cttgaaatgc tgcgtcttgc gtgactggaa 1320
gtccctaccc atcgaacggc agctgcctca cacctgctgc ggctcaggtg gaccacgccg 1380
agtcagataa gcgtcatgca acccagtttt gctttttgtg ctccagcttc cttcgttgag 1440
gagagtttga gttctctgat caggactctg cctgtcattg ctgttctctg acttcagatg 1500
aggtcacaat ctgcccctgg cttatgcagg gagtgaggcg tggtccccgg gtgtccctgt 1560
cacgtgcagg gtgagtgagg cgttgccccc aggtgtccct gtcacgtgta gggtgagtga 1620
ggcgcggccc ccgggtgtcc ctgtcccgtg cagcgtgatt gaggtgtggc ccccgggtgt 1680
ccctgtcacg tgtagggtga gtgaggcgcc atccccgggt gtccctgtca cgtgtagggt 1740
gagtgaggcg tggtccccgg gtgtccctgt cccgtgcagg gtgagtgagg cactgtcccc 1800
gggtgtccct gtcacgtgca gggtgagtga ggcgcggtcc ccgggtgtcc ctctcaggtg 1860
tagggtgagt gaggcgcggc cccagggtgt ccctgtcacg tgtagggtga gtgaggcacc 1920
gtccctgggt gtccctccca ggtatagggt gagtgaggca ctgtccccgg gtgtccctgt 1980
cacgtgcagg gtgagtgagg cgcggccccc gggtgtccct ctcaggtgca gggtgagtga 2040
ggcgctgtcc ctgggtgtcc ctgtctcgtg tagggtgagt gaggctctgt ccccaggtgt 2100
ccttggcgtt tgctcacttg agcttgctcc tgaatgtttg ctcttttctat agccacagct 2160
gcgccggttg cccattgcct gggtagatgg tgcaggcgca gtgctggtcc ccaagcctat 2220
cttttctgat gctcggctct tcttggtcac ctctccgttc cattttgcta cggggacacg 2280
ggactgcagg ctctcgcctc ccgcgtgcca ggcactgcag ccacagcttc aggtccgctt 2340
gcctctgttg ggcctggctt gctcaccacg tgcccgccac atgcatgctg ccaatactcc 2400
tctcccagct tgtctcatgc cgaggctgga ctctgggctg cctgtgtctg ctgccacgtg 2460
ttgctggaga catcccagaa agggttctct gtgccctgaa ggaaagcaag tcaccccagc 2520
cccctcactt gtcctgtttt ctcccaagct gcccctctgc ttggcccccc tgggtgggtg 2580
gcaacgcttg tcaccttatt ctgggcacct gccgctcatt gcttaggctg ggctctgcct 2640
ccagtcgccc cctcacatgg attgacgtcc agccacaggt tggagtgtct ctgtctgtct 2700
cctgctctga gacccacgtg gagggccggt gtctccgcca gccttcgtca gacttccctc 2760
ttgggtctta gttttgaatt tcactgattt acctctgacg tttctatctc tccattgtat 2820
gctttttctt ggtttattct ttcattcctt ttctagcttc ttagtttagt catgcctttc 2880
cctctaagtg ctgccttacc tgcaccctgt gttttgatgt gaagtaatct caacatcagc 2940
cactttcaag tgttcttaaa atacttcaaa gtgttaatac ttctttttaag tattcttatt 3000
ctgtgatttt tttctttgtg cacgctgtgt tttgacgtga aatcattttg atatcagtga 3060
cttttaagta ttctttagct tattctgtga tttctttgag cagtgagtta tttgaacact 3120
gtttatgttc aagatatgta gagtatcaag atacgtagag tattttaagt tatcatttta 3180
ttattgattt ctaactcagt tgtgtagtgg tctgtataat accaattatt tgaagtttgc 3240
ggagccttgc tttgtgatct agtgtgtgca tggtttccag aactgtccat tgtaaatttg 3300
acatcctgtc aatagtgggc atgcatgttc actatatcca gcttattaag gtccagtgca 3360
aagcttctgt ctcctttctag atgcatgaaa ttccaagaag gaggccatag tccctcacct 3420
gggggatggg tctgttcatt tctttctcgtt tggtagcatt tatgtgaggc attgttaggt 3480
gcatgcacgt ggtagaattt ttatcttcct gatgagtgaa tcttttggag acttctatgt 3540

ctctagtaat ctagtaattc ttttttaaa ttgctcttag tactgccaca ctgggcttct 3600
tttgattagt attttcctgc tgtgtctgtt ttctgccttt aatttatata tatatatata 3660
ttttttttt ttttgagaca gagtcttggt ctgtcgccca gggtgagtgc agtggtgtga 3720
tcacaggtca gtgtaacttt taccttctgg cctgagccgt cctctcacct cagcctcctg 3780
agtagctgga actgcagaca cgcaccgcta cacctggcta atttttaaat tttttctgga 3840
gacagggtct tgctgtgttg cccaggctgg tctcaaactc ttggactcaa gggatccatc 3900
tacctcggct tcccaaagtg ctgaattaca ggcatgagcc accatgtctg gcctaatttt 3960
caacactttt atattcttat agtgtgggta tgtcctgtta acagcatgta ggtgaatttc 4020
caatccagtc tgacagtcgt tgtttaactg gataacctga tttattttca ttttttgtc 4080
actagagacc cgcctggtgc actctgattc tccacttgcc tgttgcatgt cctcgttccc 4140
ttgtttctca ccacctcttg ggttgccatg tgcgtttcct gccgagtgtg tgttgatcct 4200
ctcgttgcct cctggtcact gggcatttgc ttttatttct ctttgcttag tgttaccccc 4260
tgatcttttt attgtcgttg tttgcttttg tttattgaga cagtctcact ctgtcaccca 4320
ggctggagtg taatggcaca atctcggctc actgcaacct ctgcctcctc ggttcaagca 4380
gttctcattc ctcaacctca tgagtagctg ggattacagg cgcccaccac cacgcctggc 4440
taatttttgt atttttagta gagataggct ttcaccatgt tggccaggct ggtctcaaac 4500
tcctgacctc aagtgatctg cccgccttgg cctcccacag tgctgggatt acaggtgcaa 4560
gccaccgtgc ccggcatacc ttgatctttt aaaatgaagt ctgaaacatt gctacccttg 4620
tcctgagcaa taagaccctt agtgtatttt agctctggcc acccccagc ctgtgtgctg 4680
ttttccctgc tgacttagtt ctatctcagg catcttgaca cccccacaag ctaagcatta 4740
ttaatattgt tttccgtgtt gagtgtttct gtagctttgc ccccgccctg cttttcctcc 4800
tttgttcccc gtctgtcttc tgtctcaggc ccgccgtctg gggtcccctt ccttgtcctt 4860
tgcgtggttc ttctgtcttg ttattgctgg taaaccccag ctttacctgt gctggcctcc 4920
atggcatcta gcgacgtccg gggacctctg cttatgatgc acagatgaag atgtggagac 4980
tcacgaggag ggcggtcatc ttggcccgtg agtgtctgga gcaccacgtg gccagcgttc 5040
cttagccagt gagtgacagc aacgtccgct cggcctgggt tcagcctgga aaaccccagg 5100
catgtcgggg tctggtggct ccgcggtgtc gagtttgaaa tcgcgcaaac ctgcggtgtg 5160
gcgccagctc tgacggtgct gcctggcggg ggagtgtctg cttcctccct tctgcttggg 5220
aaccaggaca aaggatgagg ctccgagccg ttgtcgccca acaggagcat gacgtgagcc 5280
atgtggataa ttttaaaatt tctaggctgg gcgcggtggc tcacgcctgt aatcccagca 5340
ctttgggagg ccaaggcggg tggatcacga ggtcaggagg tcgagaccat cctggccaac 5400
atgatgaaac cccatctgta ctaaaaacac aaaaattagc tgggcgtggt ggcgggtgcc 5460
tgtaatccca gctactcggg aggctgaggc aggagaattg cttgaacctg ggagttggaa 5520
gttgcagtga gccgacattg caccactgca ctccagcctg gcaacacagc gagactctgt 5580
ctcaaaaaaa aaaaaaaaaa aaaaaaaaaa aattctagta gccacattaa aaaagtaaaa 5640
aagaaaaggt gaaattaatg taataataga ttttactgaa gcccagcatg tccacacctc 5700
atcattttag ggtgttattg gtgggagcat cactcacagg acatttgaca ttttttgagc 5760
tttgtctgcg ggatcccgtg tgtaggtccc gtgcgtggcc atctcggcct ggacctgctg 5820
ggcttcccat ggccatggct gttgtaccag atggtgcagg tccgggatga ggtcgccagg 5880
ccctcagtga gctggatgtg cagtgtccgg atggtgcacg tctgggatga ggtcgccagg 5940
ccctgctgtg agctggatgt gtggtgtctg gatggtgcag gtcaggggtg aggtctccag 6000
gccctcggtg agctggaggt atggagtccg gatgatgcag gtccggggtg aggtcgccag 6060
gccctgctgt gagctggatg tgtggtgtct ggatggtgca ggtccggggt gaggtctcca 6120
ggccctcggt aagctggagg tatggagtcc ggatgatgca ggtccggggt gaggtcgcca 6180
ggccctgctg tgagctggat gtgtggtgtc tggatggtgc aggtctgggg tgaggtcacc 6240
aggccctgcg gtgagctggg tgtgcggtgt ctggatggtg caggtctgga gtgaggtcgc 6300
cagacggtgc cagaccatgc ggtgagctgg atatgcggtg tccggatggt gcaggtctgg 6360
ggtgaggttg ccaggccctg ctgtgagttg gatgtggggt gtccggatgc tgcaggtccg 6420
gtgtgaggtc accaggccct gctgtgagct ggatgtgtgg tgtctggatg gtgcaggtct 6480
ggggtgaagg tcgccaggcc cctgcttgtg agctggatgt gtggtgtctg gatggtgcag 6540
gtctggagtg aggtcgccag gccctcggtg agctggatgt gcagtgtcca gatggtgcag 6600
gtccggggtg aggtcgccag accctgcggt gagctggatg tgcggtgtct ggatggtgca 6660
ggtctggagt gaggtcgcca ggccctcggt gagctggatg tatggagtcc ggatggtgcc 6720
ggtccggggt gaggtcgcca gaccctgctg tgagctggat gtgcggtgtc tggatggtac 6780
aggtctggag tgaggtcgcc agaccctgct gtgagctgga tatgcggtgt ccggatggtg 6840
caggtcaggg gtgaggtctc caggccctcg gtgagctgga ggtatggagt ccggatgatg 6900
caggtccggg gtgaggtcgc caggccctgc tgtgaactgg atgtgcggcg tctggatggt 6960
gcaggtctgg ggtgtggtcg ccaggccctc ggtgagctgg aggtatggag tccggatgat 7020
gcaggtccgg ggtgaggtcg ccaggccctg ctgtgagctg gatgtgcggc gtctggatgg 7080
tgcaggtctg gggtgtggtc gccaggccct cggtgagctg gaggtatgga gtccggatga 7140
tgcaggtccg gggtgaggtt gccaggccct gctgtgagct ggatgtgctg tatccggatg 7200
gtgcagtccg gggtgaggtc gccaggccct gctgtgagct ggatgtgctg tatccggatg 7260
gtgcaggtct ggggtgaggt caccaggccc tgcggtgagc tggttgtgcg gtgtccggtt 7320
gctgcaggtc cggggtgagt tcgccaggcc ctcggtgagc tggatgtgcg gtgtccccgt 7380
gtccggatgg tgcaggtcca gggtgaggtc gctaggccct tggtgggctg gatgtgccgt 7440

gtccggatgg tgcaggtctg gggtgaggtc gccaggcctt tggtgagctg gatgtgcggt 7500
gtctgcatgg tgcaggtctg gggtgaggtc gccaggccct tggtgggctg gatgtgtggt 7560
gtccggatgg tgcaggtccg gcgtgaggtc gccaggccct gctgtgagct ggatgtgcgg 7620
tgtctggatg gtgcaggtcc ggggtgaggt agccaaggcc ttcggtgagc tggatgtggg 7680
gtgtccggat ggtgcaggtc cggggtgagg tcgccaggcc ctgcggttag ctggatatgc 7740
ggtgtccgga tggtgcaggt ccggggtgag gtcaccaggc cctgcggtta gctggatgtg 7800
cggtgtctgg atggtgcagg tccggggtga ggtcgccagg ccctgctgtg agctggatgt 7860
gctgtatccg gatggtgcag gtccggggtg aggtcgccag gccctgcagt gagctggatg 7920
tgctgtatcc ggatggtgca ggtctggcgt gaggtcgcca ggccctgcgg ttagctggat 7980
atgcggtgtc ggatggtgca ggtccggggt gaggtcacca ggccctgcgg ttagctggat 8040
gtgcggtgtc cggatggtgc aggtctgggg tgaggtcgcc aggccctgct gtgagctgga 8100
tgtgctgtat ccggatggtg caggtccggg gtgaggtcgc caggccctgc ggtgagctgg 8160
atgtgctgta tccggatggt gcaggtctgg cgtgaggtcg ccaggccctg cggtgagctg 8220
gatgtgcagt gtacggatgg tgcaggtccg gggtgaggtc gccaggccct gcggtgggct 8280
gtatgtgtgt tgtctggatg gtgcaggtcc ggggtgagtt cgccaggccc tgcggtgagc 8340
tggatgtgtg gtgtctggat gctgcaggtc cggggtgagt tcgccaggcc ctcggtgagc 8400
tggatatgcg gtgtccccgt gtccgaatgg tgcaggtcca gggtgaggtc gccaggccct 8460
tggtgggctg gatgtgccgt gtccggatgg tgcaggtctg gggtgaggtc gccaggccct 8520
tggtgagctg gatgtgcggt gtccggatgg tgcaggtccg gggtgaggtc accaggccct 8580
cggtgatctg gatgtggcat gtccttctcg tttaag 8616

<210> 6
<211> 2089
<212> DNA
<213> Homo sapiens

<400> 6
gtactgtatc cccacgccag gcctctgctt ctcgaagtcc tggaacacca gcccggcctc 60
agcatgcgcc tgtctccact tgcctgtgct tccctggctg tgcagctctg ggctgggagc 120
caggggcccc gtcacaggcc tggtccaagt ggattctgtg caaggctctg actgcctgga 180
gctcacgttc tcttacttgt aaaatcagga gtttgtgcca agtggtctct agggtttgta 240
aagcagaagg gatttaaatt agatggaaac actaccacta gcctccttgc ctttccctgg 300
gatgtgggtc tgattctctc tctcttttct ttttctttct tgagatggag tctcactctg 360
ttgcccaggc tggagtgcag tggcataatc ttggctcact gcaacctcca cctcctgggt 420
ttaagcgatt caccagcctc agcctcctaa gtagctggga ttacaggcac ctgccaccac 480
gcctggctaa tttttgtact tttaggagag acggggtttc accatgttgg ccaggctggt 540
ctcgaactca tgacctcagg tgatccaccc accttggcct cccaaagtgc tgggtttaca 600
ggctaagcca ccgtgcccag cccccgattc tcttttaatt catgctgttc tgtatgaatc 660
ttcaatctat tggatttagg tcatgagagg ataaaatccc acccacttgg cgactcactg 720
cagggagcac ctgtgcaggg agcacctggg gataggagag ttccaccatg agctaacttc 780
taggtggctg catttgaatg gctgtgagat tttgtctgca atgttcggct gatgagagtg 840
tgagattgtg acagattcaa gctggatttg catcagtgag ggacgggagc gctggtctgg 900
gagatgccag cctggctgag cccaggccat ggtattagct tctccgtgtc ccgcccaggc 960
tgactgtgga gggctttagt cagaagatca gggcttcccc agctcccctg cacactcgag 1020
tccctggggg gccttgtgac accccatgcc ccaaatcagg atgtctgcag agggagctgg 1080
cagcagacct cgtcagaggt aacacagcct ctgggctggg gaccccgacg tggtgctggg 1140
gccatttcct tgcatctggg ggagggtcag ggctttccct gtgggaacaa gttaatacac 1200
aatgcacctt acttagactt tacacgtatt taatggtgtg cgacccaaca tggtcatttg 1260
accagtattt tggaaagaat ttaattgggg tgaccggaag gagcagacag acgtggtggt 1320
ccccaagatg ctccttgtca ctactgggac tgttgttctg cctggggggc cttggaggcc 1380
cctcctccct ggacagggta ccgtgccttt tctactctgc tgggcctgcg gcctgcggtc 1440
agggcaccag ctccggagca cccgcggccc cagtgtccac ggagtgccag gctgtcagcc 1500
acagatgccc aggtccaggt gtggccgctc cagcccccgt gccccccatgg gtggttttgg 1560
gggaaaaggc caagggcaga ggtgtcagga gactggtggg ctcatgagag ctgattctgc 1620
tccttggctg agctgccctg agcagcctct cccgccctct ccatctgaag ggatgtggct 1680
ctttctacct gggggtcctg cctggggcca gccttgggct accccagtgg ctgtaccaga 1740
gggacaggca tcctgtgtgg aggggcatgg gttcacgtgg ccccagatgc agcctgggac 1800
caggctccct ggtgctgatg gtgggacagt caccctgggg gttgaccgcc ggactgggcg 1860
tccccagggt tgactatagg accaggtgtc caggtgccct gcaagtagag gggctctcag 1920
aggcgtctgg ctggcatggg tggacgtggc cccgggcatg gccttcagcg tgtgctgccg 1980
tgggtgccct gagccctcac tgagtcggtg ggggcttgtg gcttcccgtg agcttccccc 2040
tagtctgttg tctggctgag caagcctcct gaggggctct ctattgcag 2089

<210> 7
<211> 687
<212> DNA
<213> Homo sapiens

<400> 7
gtggctgtgc tttggtttaa cttccttttt aaacagaagt gcgtttgagc cccacatttg 60
gtatcagctt agatgaaggg cccggaggag gggccacggg acacagccag ggccatggca 120
cggcgccaac ccatttgtgc gcacagtgag gtggccgagg tgccggtgcc tccagaaaag 180
cagcgtgggg gtgtaggggg agctcctggg gcagggacag gctctgagga ccacaagaag 240
cagccgggcc agggcctgga tgcagcacgg cccgaggtcc tggatccgtg tcctgctgtg 300
gtgcgcagcc tccgtgcgct tccgcttacg gggccgggg accaggccac gactgccagg 360
agcccaccgg gctctgagga tcctggacct tgccccacgg ctcctgcacc ccacccctgt 420
ggctgcggtg gctgcggtga ccccgtcatc tgaggagagt gtggggtgag gtggacagag 480
gtgtggcatg aggatcccgt gtgcaacaca catgcggcca ggaacccgtt tcaaacaggg 540
tctgaggaag ctgggagggg ttctaggtcc cgggtctggg tggctgggga cactggggag 600
gggctgcttc tcccctgggt ccctatggtg gggtgggcac ttggccggat ccactttcct 660
gactgtctcc catgctgtcc ccgccag 687

<210> 8
<211> 494
<212> DNA
<213> Homo sapiens

<400> 8
gtgggtgccg gggacccccg tgagcagccc tgctggacct tgggagtggc tgcctgattg 60
gcacctcatg ttgggtggag gaggtactcc tgggtgggcc gcagggagtg caggtgaccc 120
tgtcactgtt gaggacacac ctggcaccta gggtggaggc cttcagcctt tcctgcagca 180
catggggccg actgtgcacc ctgactgccc gggctcctat tcccaaggag ggtcccactg 240
gattccagtt tccgtcagag aaggaaccgc aacggctcag ccaccaggcc ccggtgcctt 300
gcaccccagt cctgagccag gggtctcctg tcctgaggct cagagagggg acacagcccg 360
ccctgccctt ggggtctgga gtggtggggg tcagagagag agtgggggac accgccaggc 420
caggccctga gggcagaggt gatgtctgag tttctgcgtg gccactgtca gtctcctcgc 480
ctccactcac acag 494

<210> 9
<211> 865
<212> DNA
<213> Homo sapiens

<400> 9
gtaaggttca cgtgtgatag tcgtgtccag gatgtgtgtc tctgggatat gaatgtgtct 60
agaatgcagt cgtgtctgtg atgcgtttct gtggtggagg tacttccatg atttacacat 120
ctgtgatatg cgtgtgtggc acgtgtgtgt cgtggtgcat gtatctgtgg cgtgcatatt 180
tgtggtgtgt gtgtgtgtgg cacgtgtgtg tccatggtgt gtgtgcctgt ggtgtgcatg 240
tgtgtgtgtc tgtgacacgt gcatgttcat gctgtgtgct gcatgtctgt gatgtgccta 300
tttgtggtgt gtgtgtgcat gtgtccgtga catatgcgtg tctatggcat gggtgtgtgt 360
ggccccttgg ccttactcct tcctcctcca ggcatggtcc gcaccattgt cctcacgctc 420
tcgggtgctg gtttggggag ctccacattc agggtcctca cttctagcat gggtgccccc 480
gtcctgtcac agggctgggc cttggagact gtaagccagg tttgagagga gagtagggat 540
gctggtggta ccttcctgga cccctggcac ccccaggacc ccagtctggc ctatgccggc 600
tccatgagat ataggaaggc tgattcaggc ctcgctcccc gggacacact cctcccagag 660
cggccggggg ccttggggct cggcaggggt gaaaggggcc ctgggcttgg gttcccaccc 720
agtggtcatg agcacgctgg aggggtaagc cctcaaagtc gtgccaggcc ggggtgcaga 780
ggtgaagaag tatccctgga gcttcggtct ggggagaggc acatgtggaa acccacaagg 840
acctctttct ctgacttctt gagct 865

<210> 10
<211> 3782
<212> DNA
<213> Homo sapiens

<400> 10
tgtgggattg gttttcatgt gtgggatagg tggggatctg tgggattggt ttttatgagt 60
ggggtaacac agagttcaag gcgagctttc ttcctgtagt gggtctgcag gtgctccaac 120

agctttattg aggagaccat atcttccttt gaactatggt cgggtttata gtaagtcagg 180
ggtgtggagg cctcccctgg gctccctgtt ctgtttcttc cactctgggg tcgtgtggtg 240
cctgctgtgg tgtgtggccg gtgggcaggg cttccaggcc tccttgtgtt cattggcctg 300
gatgtggccc tggctacgct ccgtccttgg aattcccctg cgagttggag gctttctttc 360
tttctttttt tctttctttt ttttttttt tgataacaga gtctcgctct tttttgccca 420
ggctggagtg gtttggcgtg atcttggctc actgcaacct gtgcttcctg agttcaagca 480
attctcttgc ctcagcctcc caagtagctg gaattatagg cgcccaccac catgctgact 540
aatttttgta attttagtag agacgaggtt tctccatgtt ggccaggctg gtctcgaact 600
cctgacctca ggtgatcctc ccacctcggc ctcccaaagt gctgggatga caggtgtgaa 660
ccgccgcgcc cggccgagac tcgcttcctg cagcttccgt gagatctgca gcgatagctg 720
cctgcagcct tggtgctgac aacctccgtt ttccttctcc aggtctcgct aggggtcttt 780
ccatttcatg actctcttca cagaagagtt tcacgtgtgc tgatttcccg gctgtttcct 840
gcgtaattgg tgtctgctgt ttatcgatgg cctccttcca tttcctttag gctttgttta 900
ttgttgtttt tccggctcct tgaaggaaaa gtttcgatta tggatgtttg aactttcttt 960
tctaaacaag catctgaagt tgccgttttc cctctaaagc agggatcccg aggcccctgg 1020
ctgtggagtg gcaccggtct ggggcctgtt aggaacccgg cgcacagcgg gaggctaggt 1080
ggggtgtggg gagccagcgt tcccgcctga gccccgcccc tctcagatca gcagtggcat 1140
gcggtgctca gaggcgcaca caccctactg agaactgtgc gtgagagggg tctagattct 1200
gtgctcctta tgggaatcta atgcctgatg atctgaggtg gaaccgtttg ctcccaaaac 1260
catccccttc cccactgctg tcctgtggaa aaatcgtctt ccacgaaacc agtccctggt 1320
accacaatgg ttggggaccc tgtgctaaag acctgcttca gcagcctctc gtcagtgttg 1380
atatattggc ttttctgtgt tgagtccaga ataattacgg atttctgtga tgctttccgc 1440
cgacctcaga cccatgggct atttgtgggc gtgttgcctg ctcctgggtt gggaagggtg 1500
caggccccat gtaccttcct gttactgcct tccaggttgg ttctcagggt tgaatcgtac 1560
tcgatgtggt tttagcccac ggccctgccg ccagctcctg ggggctgggg aacatgctga 1620
agcacagagt caccgtgcgc gtcttttgat gcctcacaag ctcgaggcct cctgtgtccg 1680
tgttagtgtg tgtcacgtgc ctgctcacat cctgtcttgg ggacgcaggg gcttagcagg 1740
tcccgtagta aatgacaagc gtcctggggg agtctgcaga ataggaggtg ggggtgccgg 1800
tctctctccc gcgtcttcag actcttctcc tgcctgtgct gtggctgcac ctgcatccct 1860
gcaatccctc cagcactggg ctggagaggc ccgggagctc gagtgccact tgtgccacgt 1920
gactgtggat ggcagtcggt cacgggggtc tgatgtgtgg tgactgtgga tggcggttgg 1980
tcacagggggt ctgatgtgtg gtgactgtgg atggcggtcg tggggtctga tgtggtgact 2040
gtggatggcg gtcgtggggt ctgatgtgtg gtgactgtgg atggcggtcg tggggtctga 2100
tgtggtgact gtggatggcg gtcgtggggt ctgatgtggt gactgtggat ggcggtcgtg 2160
gggtctgatg tggtgactgt ggatggcagt cgtggggtct gatgtgtggt gactgtggat 2220
ggcggtcgtg gggtctgatg tggtgactgt ggatggcagt cgtggggtct gatgtgtggt 2280
gactgtggat ggcggtcgtg gggtctgatg tgtggtgact gtggatggcg gtcgtggggt 2340
ctgatgtgtg gtgactgtgg atggcggtcg tggggtctga tgtgtggtga ctgtggatgg 2400
cggtcgtggg gtctgatgtg gtgactgtgg atggcggtcg tggggtctga tgtgtggtga 2460
ctgtggatgg tgatcggtca caggggtctg atgtgtggtg actgtggatg gcggtcgtgg 2520
ggtctgatgt gtggtgactg tggatggtga tcggtcacag gggtctgatg tgtggtgact 2580
gtggatggcg gtcgtggggt ctgatgtgtg gtgactgtgg atggcggttg gtcccgggggg 2640
tctgatgtgt ggtgactgtg gatggcgatc ggtcacaggg gtctgatgtg tggtgactgt 2700
ggatggcggt cgtggggtct gatgtgtggt gactgtggat ggcggtcgtg gggtctgatg 2760
tgtggtgact gtggatggcg gtcgtggggt ctgatgtggt gactgtggat ggcggtcgtg 2820
gggtctgatg tggtgactgt ggatggcggt cgtggggtct gatgtgtggt gactgtggat 2880
ggcggttggt cccgggggtc tgatgtgtgg tgactgtgga tggcggtcgt ggggtctgat 2940
gtggtgactg tggatggcag tcgtggggtc tgatgtgtgg tgactgtgga tggcggtcgt 3000
ggggtctgat gtgtggtgac tgtggatggc ggtcgtgggg tctgatgtgt ggtgactgtg 3060
gatggcggtc gtggggtctg atgtgtggtg actgtggatg gcggtcgtgg ggtctgatgt 3120
ggtgactgtg gatggcggtc gtggggtctg atgtgtggtg actgtggatg gtgatcggtc 3180
acaggggtct gatgtgtggt gactgtggat ggcggtcgtg gggtctgatg tgtggtgact 3240
gtggatggcg gtcgtggggt ctgatgtggt gactgtggat ggcggtcgtg gggtctgatg 3300
tgtggtgact gtggatggcg gtcgtagggt ctgatgtgtg gtgactgtgg atggcagtcg 3360
gtcacagggg tctgatgtgt ggtgactgtg gatggcggtc gtggggtctg atgtgtggtg 3420
actgtggatg gcggtcgtgg ggtctgatgt gtggtgactg tggatggcgg tcgtggggtc 3480
tgatgtgtgg tgactgtgga tggcggtcgt ggggtctgat gtggtgactg tggatggtga 3540
tcggtcacag gggtctgatg tgtggtagct gcaggtggag tcccaggtgt gtctgtagct 3600
actttgcgtc ctcggccccc cggcccccgt ttcccaaaca gaagcttccc aggcgctctc 3660
tgggcttcat cccgccatcg ggcttggccg caggtccaca cgtcctgatc ggaagaaaca 3720
agtgcccagc tctggccggg gcaggccaca tttgtggctc atgccctctc ctctgccggc 3780
ag 3782

<210> 11
<211> 980
<212> DNA
<213> Homo sapiens

<400> 11
gtctgggcac tgccctgcag ggttgggcac ggactcccag cagtgggtcc tcccctgggc 60
aatcactggg ctcatgaccg gacagactgt tggccctggg gggcagtggg gggaatgagc 120
tgtgatgggg gcatgatgag ctgtgtgcct tggcgaaatc tgagctgggc catgccaggc 180
tgcgacagct gctgcattca ggcacctgct cacgtttgac tgcgcggcct ctctccagtt 240
ccgcagtgcc tttgttcatg atttgctaaa tgtcttctct gccagttttg atcttgaggc 300
caaaggaaag gtgtcccccc cctttaggag ggcaggccat gtttgagccg tgtcctgccc 360
agctggcccc tcagtgctgg gtctgaggcc aaaggaaacg tgtcccccct cttaggagga 420
cgggccgtgt ttgagccacg ccccgctgag cgggcctctc agtgctgggt ctgtccacgt 480
ggccctgtgg ccctttgcag atgtggtctg tccacgtggc cctgtggctc tttgcagatg 540
cctgttagca cttgctcggc tctaggggac agtcgtgtcc accgcatgag gctcagagac 600
ctctgggcga atttccttgg ctcccagggt gggggtggag gtggcctggg ctgctgggac 660
ccagaccctg tgcccggcag ctgggcagca actcctggat cacatatgcc atccgggcca 720
cggtgggctg tgtgggtgtg agcccagctg gacccacagg tggcccagag gagacgttct 780
gtgtcacaca ctctgcctaa gcccatgtgt gtctgcagag actcggcccg gccagcccac 840
gatggccctg cattccagcc cagccccgca cttcatcaca aacactgacc ccaaaaggga 900
cggagggtct tggccacgtg gtcctgcctg tctcagcacc caccggctca ctcccatgtg 960
tctcccgtct gctttcgcag 980

<210> 12
<211> 2485
<212> DNA
<213> Homo sapiens

<400> 12
gtgagtcagg tggccaggtg ccattgccct gcgggtggct gggcgggctg gcagggcttc 60
tgctcacctc tctcctgccc cttccccact gnccttctgc ccggggccac cagagtctcc 120
ttttctggcc cccgcccccc ccggctcctg ggctgcaggc tcccgaggcc ccggaaacat 180
ggctcggctt gcggcagccg gagcggagca ggtgccacac gaggcctgga aatggcaagc 240
ggggtgtgga gttgctcctg cgtggaggac gaggggcggg gggtgtgtct gggtcaggtg 300
tgcgccgagc gtttgagcct gcagcttgtc agctccaagt tactactgac gctggacacc 360
cggctctcac acgcttgtat ctctctctcc cgatacaaaa ggattttatc cgattctcat 420
tcctgtccct gtcgtgtgac ccccgcgagg gcgcgggctc ttctctctgt gactagattt 480
cccatctgga aagtgcgggg ttgaccgtgt agtttgctcc tctcgggggg cctgtggtgg 540
ccatggggca ggcggcctgg gagagctgcc gtcacacagc cactgggtga gccacactca 600
cggtggtaga gccacagtgc ctggtgccac atcacgtcct ctggatttta agtaaaacca 660
cacacctccc ggcaggcatc tgcctgcgac cctgtgtgtg cctggggaga gtggtagcac 720
ggaggaaatt cgtgcacact caaggtcatc agcaaggtca tccgcagtca ggtggaacgt 780
ggaggcctct ctctgggatc gtctccagcg gataaaggac tgtgcacagc ttcggaagct 840
tttatttaaa aatataacta ttaattattg cattataagt aatcactaat ggtatcagca 900
attataatat ttattaaagt ataattagaa atattaagta gtacacacgt tctggaaaaa 960
cacaaattgc acatggcagc agagtgaatt ttggccgagg gacacgtgtg cacatgtgtg 1020
taagcggccc ccaggcccac agaattcgct gacaaagtca cctccccaga gaagccacca 1080
cgggcctcct tcgtggtcgt gaattttatt aagatggatc aagtcacgta ccgtccacgt 1140
gtggcagggc tttggggaat gtgaggtgat gactgcgtcc tcatgccctg acagacagga 1200
ggtgactgtg tctgtcctgt ccctaggaca cggacaggcc cgaagctcta gtccccatcg 1260
tggtccagtt tggcctctga ataaaaacgt cttcaaaacc tgttgcccca aaaactaaga 1320
acagagagag tttcccatcc catgtgctca caggggcgta tctgcttgcg ttgactcgct 1380
gggctggccg gactcctaga gttggtgcgt gtgcttctgt gcaaaaagtg cagtcctctt 1440
gcccatcact gtgatatctg caccagcaag gaaagcctct tttcttttct ttctttttt 1500
ttttttgaga cggaacgtca ctgttgtctg cctgggcttg agtgcagtgg cgcgatctca 1560
actcactgca acctccgcct cccgggttcc agcatttctc ctgcctcagc ctcccgagca 1620
gctgagatta caggcaccca ccccctgcgc ctggctaatt tttgtatttt tagtagagag 1680
gggtttttgc catgttggcc aggctggtct cgaactcctg acctcaggtg atccacccac 1740
ctcggcctcc caaagtgctg ggattacagg tgtgagccat cacgcccagc cggaaagcct 1800
ctttttaagg tgaccaccta tagcgcttcc cgaaaataac aggtcttgtt tttgcagtag 1860
gctgcaagcg tctcttagca acaggagtgg cgtcctgtgg gctctgggga tggctgaggg 1920
tcgcgtggca gccatgcctt ctgtgtgcac ctttaggttc cacggggcta ttctgctctc 1980
actgtttgtc tgaaaacgca cccttggcat ccttgtttgg agagtttctg cttctcgttg 2040
gtcatgctga aactaggggc aaggttgtat ccgttggcgc gcagcggcta catgtagggt 2100

catgagtctt tcaccgtgga caaattcctt gaaaaaaaaa aaaggagtcc ggttaagcat 2160
tcattccggg tcaagtgtct ggttctgtga ataaactcta agatttaaga aaccttaatg 2220
aaagaaaacc ttgatgattc agagcaagga tgtggtcaca cctgtggctg gatctgtttc 2280
agccgcccca gtgcatggtg agagtgggga gcagggattg tttgttcaga ggtctcatct 2340
ggtatgtttc tgaggtgttt gccggctgaa tggtagacgt gtcgtttgtg tgtatgaggt 2400
tctgtgtctg tgtgtggctc ggtttgagtg tacgcatgtc cagcacatgc cctgcccgtc 2460
tctcacctgt gtcttcccgc cccag 2485

<210> 13
<211> 1984
<212> DNA
<213> Homo sapiens

<400> 13
gtgaggcctc ctcttcccca ggggggcttg ggtgggggtt gatttgcttt tgatgcattc 60
agtgttaata ttcctggtgc tctggagacc atgactgctc tgtcttgagg aaccagacaa 120
ggttgcagcc ccttcttggt atgaagccgc acgggagggg ttgcacagcc tgaggactgc 180
gggctccacg caggctctgt ccagcggcca tgtccagagg cctcagggct cagcaggcgg 240
gagggccgct gccctgcatg atgagcatgt gaattcaaca ccgaggaagc acaccagctt 300
ctgtcacgtc acccaggttc cgttagggtc cttggggaga tggggctggt gcagcctgag 360
gccccacatc tcccagcagg ccctcgacag gtggcctgga ctgggcgcct cttcagccca 420
ttgcccatcc cacttgcatg gggtctacac ccaaggacgc acacacctaa atatcgtgcc 480
aacctaatgt ggttcaactc agctggcttt tattgacagc agttactttt ttttttttaa 540
tactttaagt tctagggtac atgtgcacga cgtgcaggtt agttacatat gtatacatgt 600
gccatgttgg tgtgctgcac ccattaactc atcatttaca ttaggtatat ctcctaatgc 660
tatccctccc cactcccccc atcccatgac aggccctggt gtgtgatgtt ccccaccctg 720
tgtccaagtg ttctcattgt tcagttccca cctgtgagtg agaacatgtg gtgtttggtt 780
ttctttcctt gcaatagttt gctcagagtg atggtttcca gcttcgtcca tgtccctaca 840
aaggacatga actcatcctt ttttatgact gcatagtatt ccgtggtgta tatgtgccac 900
attttcttaa tccagtctat catcgatgga catttgggtt ggttgcaagt ctttgctact 960
gtgaatagtg ccgcaataaa catacgtgtg catgtgtctt tatagcagca tgatttataa 1020
tcctttgggt atatacccag taatgggatg gctgggtcaa atggtatttc tagttctaga 1080
tccttgagga atcaccacac tgtcttccac aatggttgaa ctagtttaca ctcccaccaa 1140
cagtgtaaaa gtgttctggt gctggagagg atgtggacag cagttatttt tttatgaaaa 1200
tagtatcact gaacaagcag acagttagtg aaggatgcgt caggaagcct gcaggccaca 1260
cagccatttc tctcgaagac tccgggtttt tcctgtgcat cttttgaaac tctagctcca 1320
attatagcat gtacagtgga tcaaggttct tcttcattaa ggttcaagtt ctagattgaa 1380
ataagtttat gtaacagaaa caaaaatttc ttgtacacac aacttgctct gggatttgga 1440
ggaaagtgtc ctcgagctgg cggcacactg gtcagccctc tgggacagga tacctctggc 1500
ccatggtcat ggggcgctgg gcttgggcct gagggtcaca cagtgcacca tgcccagctt 1560
cctgtggata ggatctgggt ctcggatcat gctgaggacc acagctgcca tgctggtaaa 1620
gggcaccacg tggctcagag ggggcgaggt tcccagcccc agctttctta ccgtcttcag 1680
ttatttttcc ctaagagtct gagaagtggg gccgcgcctg atggccttcg ttcgtcttca 1740
gctggcacag aattgcacaa gctgatggta aacactgagt acttataatg aatgaggaat 1800
tgctgtagca gttaactgta gagagctcgt ctgttggaaa gaaatttaag tttttcattt 1860
aaccgctttg gagaatgtta ctttatttat ggctgtgtaa attgtttgac attcagtccc 1920
tcgtagacag atactacgta aaaagtgtaa agttaacctt gctgtgtatt ttcccttatt 1980
ttag 1984

<210> 14
<211> 1871
<212> DNA
<213> Homo sapiens

<400> 14
gtgaggcccg tgccgtgtgt ctgtggggac ctccacagcc tgtgggcttt gcagttgagc 60
cccccgtgtc ctgcccctgg caccgcagcg ttgtctctgc caagtcctct ctctctgccg 120
gtgctggatc cgcaagagca gaggcgcttg gccgtgcacc caggcctggg ggcgcagggg 180
caccttcggg agggagtggg taccgtgcag gccctggtcc tgcagagacg cacccaggtt 240
acacacgtgg tgagtgcagg cggtgacctg gctcctgctg ctctttggaa agtcaagagt 300
ggcggctcct ggggccccag tgagacccca aggagctgtg cacagggcct gcagggccga 360
ggcggcagcc tcctccccag ggtgcacctg agcctgcgga gagcaggagc tgctgagtga 420
gctggcccac agcgttcgct gcggtcacgt tcctgcgtgg ggttgtttgg gatcggtggg 480
agaatttgga tttgctgagt gctgctgtct tgaaccacgg agatggctag gagtgggttt 540
cagagttgat ttttgtgaat caaactaaaa tcaggcacag gggacctggc ctcagcacag 600

gggattgtcc aatgtggtcc ccctcaaggg cgccccacag agccggtggg cttgttttaa 660
agtgcgattt gacgagggac gagaaacctt gaaagctgta aagggaaccc tcagaaaatg 720
tggccgccag gggtggtttc aggtgctttg ctgggctgtg tttgtgaaaa cccatttgga 780
cccgccctcc aagtccaccc tccaggtcca ccctccaggg ccgccctggg ctgggggtat 840
gcctggcgtt ccttgtgccg cagcccggag cacagcaggc tgtgcacatt taaatccact 900
aagattcact cgggggggagc ccaggtccca agcaactgag ggctcaggag tcctgaggct 960
gctgagggga cagagcagac ggggaacgct gcttctgtgt ggcaagttcc tgagggtgct 1020
ggccagggag gtggctcaga gtgtatgttg gggtcccacc gggggcagaa ctctgtctct 1080
gatgagtcgg cagccatgta acaggaaggg gtggccacag ggagctggga atgcaccagg 1140
ggagctgcgc agctggccga ggtcccaggg ccaggccaca ggaagggcag ggggacgccc 1200
ggggccacag cagaggccgc aggaagggaa ggggatgccc aggccagagc agaggctacc 1260
gggcacaggg gggctccctg agctgggtga gcgaggctca tgactcggcg agggaacctc 1320
cttgacgtga agctgacgac tggtgttgcc cagctcacag cccagccagg tcccgcgcct 1380
gagcaggaac tcagaaccct ccccttttgtc taaagcacag cagatgcctt cagggcatct 1440
aggagaaaac aggcaaagtc gttgagaaac gtcttaaaag aaggtgggat ggtggcaatt 1500
tcttgtccag attttagtct gccccggacc acagatgagt ctataacggg attgtggtgt 1560
tgccatgggg acacatgaga tggaccatca cagaggccac tggggctgca cctcccatct 1620
gagtcctggc tgtcccgggt ccaggccagg ttcttgcatg ctcacctacc tgtcctgccc 1680
gggagacagg gaaagcaccc cgaagtctgg agcagggctg ggtccaggct cctcagagct 1740
cctgccaggc ccagcaccct gctccaaatc accacttctc tggggttttc caaagcattt 1800
aacaagggtg tcaggttacc tcctgggtga cggccccgca tcctgggggct gacattgccc 1860
ctctgcctta g 1871

<210> 15
<211> 3801
<212> DNA
<213> Homo sapiens

<400> 15
gtgagcgcac ctggccggaa gtggagcctg tgcccggctg gggcaggtgc tgctgcaggg 60
ccgttgcgtc cacctctgct tccgtgtggg gcaggcgact gccaatccca aagggtcaga 120
ggccacaggg tgcccctcgt cccatctggg gctgagcaga aatgcatctt tctgtgggag 180
tgagggtgct cacaacggga gcagttttct gtgctatttt ggtaaaagga aatggtgcac 240
cagacctggg tgcactgagg tgtcttcaga aagcagtctg gatccgaacc caagacgccc 300
gggccctgct gggcgtgagt ctctcaaacc cgaacacagg ggccctgctg ggcatgagtc 360
cctctgaacc cgagaccctg gggccctgct gggcgtgagt ctctccgaac ccagagactt 420
cagggccctt ttgggcgtga gtctctccgc tgtgagcccc acactccaag gctcatccac 480
agtctacagg atgccatgag ttcatgatca cgtgtgaccc atcaggggac agggccatgg 540
tgtgggggggg gtctctacaa aattctgggg tcttgtttcc ccagagcccg agagctcaag 600
gccccgtctc aggctcagac acaaatgaat tgaagatgga cacagatgca gaaatctgtg 660
ctgtttcttt tatgaataaa aagtatcaac attccaggca gggcaaggtg gctcacacct 720
ataatcccag cactttggga ggccgaggtg ggtggatcac ttgaggccag gagtttgagg 780
ccaacctaac caacatagtg aaattccatt tctacttaaa aaatacaaaa attagcctgg 840
cctggtggca cacgcctgta gtccccgcta tgcgggaggc tgaggcagga gaatcatttg 900
aacccaggag gcagaggttg cagtgagccg agatcacacc actgcactcc agcctgggca 960
acagagtgag acttcatctt aaaaaaaaaa aaaaaagtat cagcattcca aaaccatagt 1020
ggacaggtgt ttttttattc tgtccttcga taatatttac tggtgctgtg ctagaggccg 1080
gaactggggg tgccttcctc tgaaaggcac accttcatgg gaagagaaat aagtggtgaa 1140
tggttgttaa accagaggtt taaactgggg tcctgtcgtt ctgagttaac agtccagatc 1200
tggactttgc ctctttccag aatgctccct ggggtttgct tcatgggggga gcagcaggtg 1260
tggacaccct cgtgatgggg gagcagcagg tgcagacgcc ctcatgatgg gggagtggca 1320
ggtgcagaca cccttgtgca tggtgcccag catgtccctg ttgcagctcc ctccccacaa 1380
ggatgccggt ctcctgtgct ccccacagtc cctgcttccc tctcacagcc ttacctggtc 1440
ctggcctcca ctggctttgt ctgcatgatt tccacatttc ctgggctccc agcacctctt 1500
cgcctctccc aggcacctct gcagtgctgg ccataccagt cagctgtgaa ctgtccactg 1560
cttattttgc tccccatgaa atgtattttt taggacaggc acccctggtt ccagcctctg 1620
gcacagcatc agtgaatgtt attgaaggac aaaggacaga caaacaaatc aggaaaatgg 1680
gttctctcta aacacattgc aaagccacag aggctagtgc aggatgggtg ggcatcaggt 1740
catcagatgt gggtccaatg ccagaatatt ctgtgctccc aaaggccact tggtcagagt 1800
gtgtgcttgc agaggtggct ctaaaagctc agcagtggag gcagtggttc gccatactca 1860
gggtgaactc acatcctctg tgtctgaagt atacagcaga ggcttgaagg gcatctggga 1920
gaagaaaaca ggcaaaatga ttaagaaaag tgaaaaagga aaagtggtaa gatgggaatt 1980
ttcttgtcca gattttagtc tcccaaacca cagctcagat ggtagaatgt ggtcagaact 2040
gatggacaga acaatagaac aaaacggaag ccctatctct cagaaacgtg tgttaatgtg 2100
gtatgtggca cagctgatgg aaaagagagt gtgtgtgtaa ttttttttc tgagaaaact 2160

gactggaagc aaataagttg tgtctttaca gcatatacca gagcagattc taggtagaag 2220
aggagacaca tgcaaacaac accagcaaca gaaataaaac aaaagactca aagggaaggg 2280
aggtgaacgt tccctggttt ggtgttgggg aaggacacac agggaggcgg atgaaaccag 2340
tgaggcaacg ggcattgctt tcactgcaga gaaactcagc ttgcctgagc cacagtgaaa 2400
atggccattc cctggagcgt ttgtgcacgt gatttattta aggcgccctg tgaggtcctg 2460
cacattcatc ctctcacttt gttctcctaa ccacctgaga ggtagaggag gaaaggctcc 2520
aggggagcag ccgcccttgg tcacccagct ggcaaagggc atgcatgatt gcagcctggc 2580
ctcctgctcc ggggcccttg ctctgcccga ggaccccaca caagtcagac ccataggctc 2640
agggtgagcc ggagcccaag gtcgtgttgg ggatggctgt gaaagaagaa atggacgtct 2700
gatgcacact tgggaaggtc ctaccagcag cgtcaaagaa atgcatgtga aactgacagc 2760
gagacccatc cctcaaagaa acgcacgtga aactgatggc gagacctgtc cccatccctc 2820
atgctggctc cttttctggg cttgccaaga gccagcatca ggttgaggca agctggaaag 2880
acttttctgg aaagcagctt gtttgcatgg aagtcctcac aatgtcctgt gtcttcccag 2940
taattccact tctgaagtga ccagacatta tcacgggtct tatttaccat ttccagtgtt 3000
ccaggcaggg ggacttgcca cagcaagtca cgaacctgcc caaatacagg gctaaggaga 3060
tattatgcat cacaaaactt gctctgccat taaacatttt tcaaagaatt tttgaagaat 3120
gtttaatggc acaaaacgtt tatttcaatg tagcagtgtt caaagctgga tgtaaaagaa 3180
cacaccccag gagcctgccg tgaatgtcat gtgtgttcat ctttggacat ggacatacat 3240
gggcagtgag tggtggtgag gccctggagg acatcggtgg gatgcctcca tcctgccccc 3300
ctggagacac catgtgtgcc acgtgcactc actggagccc tgtttagctg gtgccacctg 3360
gctcttccat ccctgagatt caaacacagt gagattcccc acgcccaact cagtgttctc 3420
ccacaaaaaa cctgagtcac acctgtgttc actcgaggga cgcccgggag ccagggctcc 3480
acagtttatt atgtgttttt ggctgagtta tgtgcagatc tcatcagggc agatgatgag 3540
tgcacaaaca cggccgtgcg aggtttggat acactcaaca tcactagcca ggtcctggtg 3600
gagtttggtc atgcagagtc tggatggcat gtagcatttg gagtccatgg agtgagcacc 3660
cagccccctc gggctgcagc gcatgcccca ggcaggacaa ggaagcggga ggaaggcagg 3720
aggctctttg gagcaagctt tgcaggaggg ggctgggtgt ggggcaggca cctgtgtctg 3780
acattccccc ctgtgtctca g 3801

<210> 16
<211> 880
<212> DNA
<213> Homo sapiens

<400> 16
gtgagcaggc tgatggtcag cacagagttc agagttcagg aggtgtgtgc gcaagtatgt 60
gtgtgtgtgt gtgcgcgcgt gcctgcaagg ctgatggtga ctggctgcac gtaagagtgc 120
acatgtacgc atatacacgt gagcacatac atgtgtgcat gtgtgtacat gaaggcatgg 180
cagtgtgtgc acaggtgtgc aagggcacaa gtgtgtgcac atgcgaatgc acacctgaca 240
tgcatgtgtg ttcgtgcaca gtcgtgtggg cattcacgtg aggtgcatgc gtgtgggtgt 300
gcagtgtgag tagcatgtgt gcacataaca tgtattgagg ggtcctcgtg ttcaccccgc 360
taggtcctca gcaccagtgc cactccttac aggatgagac ggggtcccag gccttggtgg 420
gctgaggctc tgaagctgca gccctgaggg cattgtccca tctgggcatc cgcgtccact 480
ccctctcctg tgggcttctg tgtccactcc ccctctcctg tgggcattta catccactcc 540
actccctctc tcctgtgggc atccgcgtcc actccccctc tctgtgggca tctgcgtcca 600
cctccccctct ctgtgggcat ttgcgtccac tccctctcct ggttccttcc tgtcttggcc 660
gagcctcggg ggcaggcaga tgacacagag tcttgactcg cccagggtgg ttcgcagctg 720
ccgggtgagg gccaggccgg atttcactgg gaagagggat agtttcttgt caaaatgttc 780
ctctttcttg ttccatctga atggatgata aagcaaaaag taaaaacttа aaatcccaga 840
gaggtttcta ccgtttctca ctctttcttg gcgactctag 880

<210> 17
<211> 3186
<212> DNA
<213> Homo sapiens

<400> 17
gtgagccgcc accaaggggt gcaggcccag cctccaggga ccctccgcgc tctgctcacc 60
tctgacccgg ggcttcacct tggaactcct gggttttagg ggcaaggaat gtcttacgtt 120
ttcagtggtg ctgctgcctg tgcacagttc tgttcgcgtg gctctgtgca aagcacctgt 180
tctccatctc tgggtagtgg taggagccgg tgtggcccca ggtgtcccca ctgtgcctgt 240
gcactggccg tgggacgtca tggaggccat cccagggcag caggggcatg gggtaaagag 300
atgtttatgg ggagtcttag cagaggaggc tgggaaggtg tctgaacagt agatgggaga 360
tcagatgccc ggaggatttg gggtctcagc aaagagggcc gaggtgggtg caggtgaggg 420
tcgctggccc caccccggg aaggtgcagc agagctgtgg ctccccacac agcccggcca 480

gcacctgtgc tctgggcatg gctgtgctcc tggaacgttc cctgtcctgg ctggtcaggg 540
ggtgcccctg ccaagaatcg acaactttat cacagaggga agggccaatc tgtggaggcc 600
acagggccag cttctgcctg gagtcagggc aggtggtggc acaagcctcg gggctgtacc 660
aaagggcagt cgggcaccac aggcccgggc ctccacctca acaggcctcc cgagccactg 720
ggagctgaat gccaggaggc cgaagccctc gccccatgag ggctgagaag gagtgtgagc 780
atttgtgtta cccagggccg aggctgcgcg aattaccgtg cacacttgat gtgaaatgag 840
gtcgtcgtct atcgtggaaa cccagcaagg gctcacggga gagttttcca ttacaaggtc 900
gtaccatgaa aatggttttt aacccgagtg cttgcgcctt catgctctgg cagggagggc 960
agagccacag ctgcatgtta ccgcctttgc accagctcca gaggcttggg accaggctgt 1020
ctcagttcca gggtgcgtcc ggctcagacc gccctcctct ctgccttctc tctctgcctc 1080
aaatcttccc tcgtttgcat ctccctgacg cgtgcctggg ccctcgtgca agctgcttga 1140
ctccttccg gaaaccttg gggtgtgctg gatacaggtg ccactgagga ctggaggtgt 1200
ctgacactgt ggttgacccc agggtccagc tggcgtgctt ggggcctcct tgggccatga 1260
tgaggtcaga ggagttttcc caggtgaaaa ctcctgggaa actcccaggg ccatgtgacc 1320
tgccacctgc tcctcccata ttcagctcag tcttgtcctc atttccccac cagggtctct 1380
agctccgagg agctcccgta gagggcctgg gctcagggca gggcggctga gtttccccac 1440
ccatgtgggg acccttgggt agtcgcttga ttgggtagcc ctgaggaggc cgagatgcga 1500
tgggccacgg gccgtttcca aacacagagt caggcacgtg gaaggcccag gaatcccctt 1560
ccctcgaggc aggagtggga gaacggagag ctgggccccg atttcacggc agccaggctg 1620
cagtgggcga ggctgtggtg gtccacgtgg cgctgggggc ggggtctgat tcaaatccgc 1680
tggggctcgg ccttcctggc ccgtgctggc cgcgcctcca cacgggcttg gggtggacgc 1740
cccgacctct agcaggtggc tatttctccc tttggaagag agcccctcac ccatgctagg 1800
tgtttccctc ctgggtcagg agcgtggccg tgtggcaacc ccgggacctt aggcttattt 1860
atttgtttaa aaacattctg ggcctggctt ccgttgttgc taaatgggga aaagacatcc 1920
cacctcagca gagttactga gaggctgaaa ccggggtgct ggcttgactg gtgtgatctc 1980
aggtcattcc agaagtggct caggaagtca gtgagaccag gtacatgggg ggctcaggca 2040
gtgggtgaga tgaggtacac gggggggctca ggcagtgggt gaggccaggt acatgggggg 2100
ctcaggcact gggtgagatg aggtacacgg ggggctcagg cagagggtca gaccaggtac 2160
acgggggctc tgatcacacg cacatatgag cacatgtgca catgtgctgt ttcatggtag 2220
ccaggtctgt gcacacctgc cccaaagtcc caggaagctg agaggccaaa gatggaggct 2280
gacagggctg gcgcggtggc tcacacctgt agtcccagca ctttgggagg ccgaggcgag 2340
aggatccctt gagcccagga gtttaagacc agcctgagca acatagtaga accccatctc 2400
tatgaaaaat aaaaacaaaa attagctgaa catggtggtg tgcgcctgta gttccaatac 2460
ttgggaggct gaagtgggag gatcacttga gcccaggagg tggaagctgc agtgagctga 2520
gattgcacca ctgtactgca gcctgggtga cagagtgaga gcccatctca acaacaacaa 2580
agaagactga caaatgcagt ttcttggaaa gaaacattta gtaggaactt aacctacaca 2640
cagaagccaa gtcggtgtct cggtgtcagt gagatgagat gatgggtcct cacaccatca 2700
ccccagaccc agggtttatg caccacaggg gcgggtggct cagaagggat gcgcaggacg 2760
ttgatatacg atgacatcaa ggttgtctga cgaagggcag gattcatgat aagtacctgc 2820
tggtacacaa ggaacaatgg ataaactgga aaccttagag gccttcccgg aacaggggct 2880
aatcagaagc cagcatgggg ggctggcatc caggatggag ctgcttcagc ctccacatgc 2940
gtgttcatac agatggtgca cagaaacgca gtgtacctgt gcacacacag acacgcagct 3000
actcgcacac acaagcacac acacagacat gcatgcatgc atccgtgtgt gtgcacctgt 3060
gcccatgagg aaacccatgc atgtgcattc atgcacgcac acaggcaccg gtgggcccat 3120
gcccacaccc acgagcaccg tctgattagg aggcctttcc tctgacgctg tccgccatcc 3180
tctcag 3186

<210> 18
<211> 781
<212> DNA
<213> Homo sapiens

<400> 18
gtatgtgcag gtgcctggcc tcagtggcag cagtgcctgc ctgctggtgt tagtgtgtca 60
ggagactgag tgaatctggg cttaggaagt tcttacccct tttcgcatca ggaagtggtt 120
taacccaacc actgtcaggc tcgtctgccc gcctctcgt ggggtgagca gagcacctga 180
tggaagggac aggagctgtc tgggagctgc catccttccc accttgctct gcctggggaa 240
gcgctggggg gcctggtctc tcctgtttgc cccatggtgg gatttggggg gcctggcctc 300
tcctgtttgc cctgtggtgg gattgggctg tctcccgtcc atggcactta gggcccttgt 360
gcaaacccag gccaagggct taggaggagg ccaggcccag gctaccccac ccctctcagg 420
agcagaggcc gcgtatcacc acgacagagc cccgcgccgt cctctgcttc ccagtcaccg 480
tcctctgccc ctggacactt tgtccagcat cagggaggtt tctgatccgt ctgaaattca 540
agccatgtcg aacctgcggt cctgagctta acagcttcta ctttctgttc tttctgtgtt 600

gtggaaattt cacctggaga agccgaagaa aacatttctg tcgtgactcc tgcggtgctt 660
gggtcgggac agccagagat ggagccaccc cgcagaccgt cgggtgtggg cagctttccg 720
gtgtctcctg ggaggggagc tgggctgggc ctgtgactcc tcagcctctg ttttccccca 780
g 781

<210> 19
<211> 536
<212> DNA
<213> Homo sapiens

<400> 19
gcaagtgtgg gtggaggcca gtgcgggccc cacctgccca ggggtcatcc ttgaacgccc 60
tgtgtggggc gagcagcctc agatgctgct gaagtgcaga cgccccgggg cctgaccctg 120
ggggcctgga gccacgctgg cagccctatg tgattaaacg ctggtgtccc caggccacgg 180
agcctggcag ggtccccaac ttcttgaacc cctgcttccc atctcagggg cgatggctcc 240
ccacgcttgg gagccttctg acccctgacc tgtgtcctct cacagcctct tccctggctg 300
ctgccctgag ctcctggggt cctgagcaag ttctctcccc gccccgccgc tccagcgtca 360
ctgggctgcc tgtctgctcg ccccggtgga ggggtgtctg tcccttcact gaggttccca 420
ccagccaggg ccacgaggtg caggccctgc ctgcccggcc acccacacgt cctaggaggg 480
ttggaggatg ccacctctgg cctcttctgg aacggagtct gattttggcc ccgcag 536

<210> 20
<211> 3179
<212> DNA
<213> Homo sapiens

<400> 20
atctcatgtt tgaatcctaa tgtgcactgc atagacacca ctgtatgcaa ttacagaagc 60
ctgtgagtga acggggtggt ggtcagtgcg ggcccatggc ctggctgtgc atttacggaa 120
gtctatgagt gaatgggggtt gtggtcagtg cgggcccatg gcctggctgg gcctgggagg 180
tttctgatgc tgtgaggcag gaggggaagg agggtagggg atagacagtg ggagccccca 240
ccctggaaga cataacagta agtccaggcc cgaagggcag cagggatgct gggggcccag 300
cttgggcggc ggggatgatg gagggcctgg ccagggtggc agggatgatg ggggccccag 360
ctggggtggc aggggtgatg gggggggctg gtctgggtgg cggggaagat ggggaagcct 420
ggctgggccc cctcctcccc tgcctcccac ctgcagccgt ggatccggat gtgcttccct 480
ggtgcacatc ctctgggcca tcagctttca tggaggtggg gggcaggggc atgacaccat 540
cctgtataaa atccaggatt cctcctcctg aacgccccaa ctcaggttga aagtcacatt 600
ccgcctctgg ccattctctt aagagtagac caggattctg atctctgaag ggtgggtagg 660
gtggggcagt ggagggtgtg gacacaggag gcttcagggt ggggctggtg atgctctctc 720
atcctcttat catctcccag tctcatctct catcctctta tcatctccca gtctcatctg 780
tcttcctctt atctcccagt ctcatctgtc atcctcttac catctcccag tctcatctct 840
tatcctctta tctcctagtc tcatccagac ttacctccca gggcgggtgc caggctcgca 900
gtggagctgg acatacgtcc ttcctcaggc agaaggaact ggaaggattg cagagaacag 960
gaggggcggc tcagagggac gcagtcttgg ggtgaagaaa cagcccctcc tcagaagttg 1020
gcttgggcca cacgaaaccg agggccctgc gtgagtggct ccagagcctt ccagcaggtc 1080
cctggtgggg ccttatggta tggccgggtc ctactgagtg caccttggac agggcttctg 1140
gtttgagtgc agcccggacg tgcctggtgt cggggtgggg gcttatggcc actggatatg 1200
gcgtcattta ttgctgctgc ttcagagaat gtctgagtga ccgagcctaa tgtgtatggt 1260
gggcccaagt ccacagactg tgtcgtaaat gcactctggt gcctggagcc cccgtatagg 1320
agctgtgagg aaggaggggc tcttggcagc cggcctgggg gcgcctttgc cctgcaaact 1380
ggaagggagc ggccccgggc gccgtgggcg gacgacctca agtgagaggt tggacagaac 1440
agggcgggga cttcccagga gcagaggccg ctgctcaggc acacctgggt ttgaatcaca 1500
gaccaacagg tcaggccatt gttcagctat ccatcttcta caaagctcca gattcctgtt 1560
tctccgggtg ttttttgttg aaattttact caggattact tatatttttt gctaaagtat 1620
tagaccctta aaaaaggtat ttgctttgat atggcttaac tcactaagca cctactttat 1680
ttgtctgttt ttatttatta ttattattat tattagagat ggtgtctact ctgtcaccca 1740
ggttgttagt gcagtggcac agtcatggct cgctgtagcc gcaaaccccc aggctcaagt 1800
gatcctccgg cctcagcttc ccagagtgct gggattacag gtgtgagcca ctgcccttgc 1860
ctggcacttt taaaaaccac tatgtaaggt caggtccagt ggcttccaca cctgtcatcc 1920
cagtagtttg ggaagccgag gcagaaggat tgtctgaggc caggagtttg agaccagcat 1980
gggtaacata gggagacccc atctctacaa aaaatgcaaa aagttatccg ggcgtggggt 2040
ccagcatctg tagtcccagc tgctcgggag gctgagtggg aggatcgctt gagcccggga 2100
ggtcatggct gcagtgagct gtgattgtac catcgcactc cagcctgggc aacagagtga 2160
gaccctgtct caaaaaaaaa aaaaaaaaaa gaaggagaag gagaagagaa gaagaaggaa 2220
gaaggaaaga gaagaagaag gaagaaggaa gaaagaagga gaaggaggcc tgctaggtgc 2280

```
taggtagact gtcaaatctc agagcaaaat gaaaataaca aagttttaaa gggaaagaaa 2340
aaccccagct ctttggactt ccttaggcct gaacttcatc tcaagcagct tccttccaca 2400
gacaagcgtg tatggagcga gtgagttcaa agcagaaagg gaggagaagc aggcaagggt 2460
ggaggctgtg ggtgacacca gccaggaccc ctgaaaggga gtggttgttt tcctgcctca 2520
gccccacgct cctgccggtc ctgcacctgc tgtaaccgtc gatgttggtg ccaggtgccc 2580
acctgggaag gatgctgtgc agggggcttg ccaaactttg gtgggtttca gaagccccag 2640
gcacttgtgg caggcacaat tacagcccct ccccaaagat gcccacgtcc ttctcctgga 2700
acctgtgaat gtgtcacccg caaggcagag gctggtgaag gctgcaggtg gaatcacggc 2760
tgccagtcag ccgatcttaa ggtcatcctg gattatctgg tgggcctgat atggccacaa 2820
gggtccctag aagtgagaga gggaggcagg ggagagtcag agaggggacg tgagaaggac 2880
cactggccac tgctggcttt gagatggagg agggggtccc cagccaagga atgggggcag 2940
ccgctccatg ctggaaaagc aagcaatcct ccccggtcct gagggcacac ggccctgccc 3000
acgcctcgat ttcaggccag tgggacctgt ttcagctttc cggcctccag agctgtaaga 3060
tgatgcgttt gtgttcagcc actaagctgc agtgattcgt cacagcagca aatggaatag 3120
cagtacaggg aaatgaatac agggacagtt ctcagagtga ctctcagccc acccctggg  3179
```